DOCUMENT RESUME

ED 192 340 CS 205 854

AUTHOR Holland, V. Melissa; Rose, Andrew M.
TITLE Understanding Instructions with Complex Conditions. Technical Report No. 5.
INSTITUTION American Institutes for Research, Washington, D.C.; Carnegie-Mellon Univ., Pittsburgh, Pa.; Siegel & Gale, Inc., New York, N.Y.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO AIR-75002-2/80-TR
PUB DATE Feb 80
CONTRACT 400-78-0043
NOTE 89p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adults; *Difficulty Level; *Readability; *Reading Comprehension; Reading Processes; Reading Rate; *Reading Research; *Sentence Structure
IDENTIFIERS *Document Design Project; *Instructions

ABSTRACT

Forty adults served as subjects in a study of the comprehensibility of instructions conditioned by multicategory compounds, such as "If you are male, or both married and not employed, go to question 3." The stimulus materials were 96 sentence frames of the form, "If you are X, press (button) Y," in which "X" was replaced by coordinate compounds generated by different combinations of five factors: (1) number of categories in the antecedent statement, (2) organization of categories, (3) "and" versus "or," (4) presence or absence of negation, and (5) organizational subtypes. In individual sessions, subjects responded to all 96 instructions, presented one at a time on a computer display terminal. Analyses of subjects' response time and accuracy yielded a number of results, including that the pattern of results, for both speed and accuracy, was primarily of the interactive type and that each of the major sentence dimensions appeared to contribute to overall sentence difficulty. The results have both theoretical and practical implications and lend support to the premise that complex instructions should be experimentally evaluated prior to inclusion in a public document. (GT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192340

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AIR-75002-2/80-TR

DOCUMENT DESIGN PROJECT

Technical Report No. 5

Understanding Instructions with Complex Conditions

V. Melissa Holland
Andrew M. Rose

A product of the Document Design Project funded by the National Institute of Education, DHEW, Washington, DC
Contract No. NIE-400-78-0043

February 1980

CS205854

AMERICAN INSTITUTES FOR RESEARCH/1055 Thomas Jefferson Street, NW, Washington, DC 20007 • 202/342-5000
with Carnegie Mellon University and Siegel & Gale, Inc.

OCT 2 0 1980

This report is part of the work being done under contract #400-78-0043 of the National Institute of Education. The project officer is Dr. Ramsay Selden of the Teaching and Learning/ Reading and Language Group. This report does not necessarily reflect the views of the funding agency.

ABOUT THE DOCUMENT DESIGN PROJECT

In September 1978, the American Institutes for Research (AIR) began the Document Design Project to foster clear and simple writing and design of public documents. The purpose of the Document Design Project (DDP) is to help make forms, regulations, brochures, and other written materials easier for people to read, to understand, and to use.

Carnegie-Mellon University and Siegel & Gale, Inc., a private firm that specializes in language simplification and language training, work with AIR on this project. Funding for the project comes from the National Institute of Education.

The Project's goal is to increase the knowledge and skills of people who produce public documents. To accomplish this goal, the Document Design Project is working on three tasks:

Task 1: To conduct theoretical and applied research studies on language comprehension, on the ways in which skilled and unskilled writers work, on problems associated with different document features.

Task 2: To work directly with government and private agencies as they produce materials for public use.

Task 3: To develop courses on writing and design for graduate students and undergraduates.

If you have questions or comments on this report or on other work of the Document Design Project, contact Dr. Janice C. Redish, Director, The Document Design Project, AIR, (202) 342-5071.

ACKNOWLEDGMENTS

We would like to thank the people and organizations who helped in conducting this research. Dr. Steve Shwartz of the Johns Hopkins University Psychology Department (now of Yale University) wrote the program which enabled us to computerize the experimental procedure and scoring. Ms. Ingrid Heinsohn, a research associate at the American Institutes for Research, served as experimenter for the study. Several employees of the Close-Up Foundation in Washington, D.C. volunteered to participate as experimental subjects.

We also thank Mr. George Wheaton, Principal Research Scientist, and Dr. Janice Redish, Director of the Document Design Project--both of the American Institutes for Research--for technical and editorial comments which improved the paper.

Table of Contents

PREFACE 1

PURPOSE 3

Document Problems: Practical Motivation for Studying Complex Conditions on Instructions 3

Theoretical Motivation for the Study: The Limits of Previous Research 8

Sentence comprehension in the verification paradigm . 8

Logical rules in the concept-learning paradigm . . . 15

Reasoning processes: Evidence and theory 16

Negation 20

The nature of sentence reference in the verification paradigm 22

Partial processing in sentence comprehension strategies 23

Recapitulation 23

DESCRIPTION OF STIMULI AND SPECIFIC HYPOTHESES 24

Factor 1: Category Number 27

Predictions 27

Rationale for defining the levels 27

Factor 2: Organizational Type 28

Predictions 28

Rationale for defining organizational types 30

Factor 3: "And" vs. "Or" 30

Predictions and rationale 30

Factor 4: Affirmative-Negative 31

Predictions 32

Rationale for selection 32

Factor 5 (Nested): Organizational Subtypes 32

(a) Simple grouping 32

(b) "Not" grouping 34

(c) String subtypes: Explicit vs. implicit 35

METHOD 38
Materials 38
Description of categories 38
Rationale for choosing categories 38
Sentence generation 39
Implementation 39
Procedure 40
Design 42
Subjects 42
RESULTS AND DISCUSSION 43
Results of Analysis 45
Response Time 45
Accuracy 57
Summary and Implications 60
Practical Implications 62
APPENDIX A: ANOVA Tables for Each of the Experimental Sentence Types A-1
APPENDIX B: An Information-Processing Model B-1

LIST OF TABLES

Table 1: Combinations of Individual Category Truth Values That Make Sentence (3) True 19
Table 2: Group Size and Sequence in Simple Groupings with 2 and 3 Branches 34
Table 3: The Stimulus Set 37
Table 4: Probability of Errors by Type 58

LIST OF FIGURES

Figure 1: Response Time for Strings 46
Figure 2: Response Time for Simple Groupings 47
Figure 3: Response Time for "Not" Groupings 48
Figure 4: Percent Correct for Strings 49
Figure 5: Percent Correct for Simple Groupings . . . 50
Figure 6: Percent Correct for "Not" Groupings 51
Figure 7: Response Times and Percent Correct for Single-Category Negation 52
REFERENCES 65

PREFACE

As part of the Document Design Project at the American Institutes for Research we have conducted an empirical study of the comprehensibility of a type of instruction which appears particularly taxing to readers: instructions conditioned by multi-category compounds, such as "If you are male, or both married and not employed, go to question 3." In this report, we describe the details of this study in four parts: (I) Purpose, (II) Stimulus Description and Hypotheses, (III) Method, and (IV) Results and Discussion. The rationale for this study is both practical and theoretical, and we intend the report for both applied and academic audiences.

Part I (Purpose) is divided into two sections. The first describes the practical bases for this study. The second section reviews relevant research and theory. Part IV presents the results and discusses both theoretical and practical implications. The theoretical section refers to an explanatory model for processing multi-category compound sentences, presented in an appendix. The applied section draws from the results a set of recommendations for designing instructions.

UNDERSTANDING INSTRUCTIONS WITH COMPLEX CONDITIONS

Purpose

Here we discuss the rationale for our study of complex conditional sentences. We will look at the rationale from two perspectives: practical and theoretical.

Document Problems: Practical Motivation for Studying Complex Conditions on Instructions

Documents ranging from application forms to operating manuals commonly instruct actions ("submit the following") or define status (eligibility) subject to specific conditions, such as "Only disabled veterans are eligible." Frequently these conditions are complex, consisting of many categories related by an intricate logic: "All veterans who are either employed or disabled and either single or not over 25 are eligible." Alternatively, these conditions can be specified in question form: "Are you either employed or disabled and either single or not over 25?" In order to answer the question, apply the statement, or follow the instruction, readers must first grasp the relationship--or logical structure--that links the categories, and then decide whether the condition so defined fits their particular circumstances.

Sometimes the combination of categories becomes so complex that to interpret the sentence requires a second or third reading. The complexity of this statement, taken from an individual retirement annuity brochure, is not unusual:

> You are eligible for an IRA in any given tax year if you work for yourself or someone else, full-time or part-time, and are not an active participant during any part of the tax year in any of the following types of programs: a qualified pension, profit sharing or stock bonus plan; a qualified annuity plan; a qualified bond purchase plan...; an annuity contract...; a qualified retirement plan...; or a government retirement plan. ("Individual Retirement Account: Plan for Your Retirement," Pension Benefit Guarantee Corporation, #801.)

Such a sentence is obviously difficult, as is confirmed by comments from users and designers of this and other forms. Whether the set of conditions is expressed by the antecedent ("if") in a conditional ("if-then") statement, or by a question, or by some other construction, the problem lies not so much in the terminology as in the way the categories are grouped: the logical structure which must be grasped to interpret and apply the conditions.

Aware of such problems, specialists in written communication (e.g., Morris, 1972; Horabin & Lewis, 1974) advise rewriting or reformatting complex conditions in definitions or instructions. For example, Morris suggests: "When the [instruction] you are working with is complex and full of conditions, use a decision table .... [This is] a substantial improvement over ... a narrative."

But this advice lacks both definition and support. It does not specify what type and degree of complexity call for reformatting, nor does it provide evidence that the recommended changes will work. Typically, document design guidelines such as this are suggested with no empirical data for defining problems or prescribing solutions.

To develop guidelines that speak with validity and precision to the problem of interpreting sentences with complex conditions, we need experimental evidence for:

- diagnosing the problem: what dimensions underlie difficulties in comprehending and applying complex conditions, and at what levels do these dimensions become a problem?
- bounding the problem: what is the extent of difficulties people have?
- deriving solutions: what can be done to alleviate the difficulties by way of writing or presenting problem sentences in new ways?

Answering these questions requires resolving theoretical and practical issues. To talk about solutions, we require a rational prescription -- deriving rules for what to do from explanations of what is causing the problem. Psycholinguistic theory as currently developed is sufficiently explanatory to enable some predictions about sources of difficulty. But it is not prescriptive. Applied literature connects problems and solutions, but it is either intuitive (the writing guidelines) and lacking evidence, or ad hoc (as in practical research) and lacking explanation. Recognizing the need for a coordinated approach in formulating guidelines, we have begun a program of research into sentences with complex conditions that is designed to wed theory and application.

We are dividing this research into the three functional stages outlined above. Before investigating solutions, we are studying the prerequisite areas of problem diagnosis and bounding. Our problem domain can be defined more precisely: the psychological complexity of sentences expressing multi-category conditions linked by logical connectives.* Since neither theoretically nor practically motivated research has probed this domain, our inquiry is preliminary and exploratory.

We have begun the first stage of research with a study designed to classify, measure, and explain some of the difficulties imposed by complex conditions. To make the study manageable, we have restricted our inquiry to selected aspects of the problem:

- We chose instructions ("Do Y") rather than definitions or other kinds of communicative acts on which conditions can be placed.
- We presented conditional sentences (the "if-then" construction) rather than other forms for expressing conditions on instructions. Specifically, we integrated the imperative with the conditional form: "If X, do Y."

*By "category" we mean a term of description or reference, generally a noun or adjective. Here, categories function as predicate complements: nouns (you are a veteran) or adjectives (you are employed).

- We required personal referents for conditions instead of general referents: "If you are X, do Y."
- We manipulated structural dimensions of conditions while ignoring semantics; that is, we looked at how categories combine rather than what they contain.
- We used simple, over-learned, one-word categories (e,g., "male," "female").
- We limited the manner in which the antecedent categories combine to include coordinate relations ("and" or "or") and exclude subordinates (e.g., "after," "whom," or "belonging to").

So designed, the study can be characterized as a search for structural determinants of complexity in conditional antecedents, targeting structures in which several categories are joined coordinately ("If you are A, B, C, and D, do X."). We have chosen structural variables based on well-established correlates of sentence difficulty, combined with the results of a pilot study specifically of multi-category conditional sentences. The pilot results suggested that, besides number of categories, other sentence features make a considerable difference for performance: the type of connective, the presence of subconditions (groupings of categories), the presence and scope of negation, and the potential ambiguity of strings of commas. Most of these dimensions can be considered aspects of grammatical as well as logical structure. Negation is a phenomenon both of syntax and reasoning; and the last dimension, punctuation, is purely surface syntax. However, we have concentrated on the logical rule relating categories, defined by the type of connective and the presence or absence of subconditions.

On the performance side, we are concerned with how these dimensions affect two in a sequence of three theoretical steps assumed to underlie processing of conditional instructions:

(1) understanding the conditional clause in the antecedent (in psycholinguistic terms, extracting the logical rule along with the category meanings and encoding these in an internal representation of the clause);

(2) evaluating the truth of the condition (comparing a representation of the antecedent with a representation of personal characteristics to decide whether the former applies); and

(3) using the evaluation decision to determine a response to the imperative clause.

We assume that understanding and evaluating are the primary processes affected by conditional structure. We have tested these effects by measuring the speed and accuracy with which subjects perform conditional instructions. The basic task is a modified sentence verification procedure.*

To illustrate how the general questions guiding this research apply to the specific structural effects we have hypothesized, we can look at the variable of number of categories. To diagnose and bound comprehension problems in terms of this variable, we could ask:

- How many categories can subjects handle before errors, re-reading, or other problem indications emerge?
- How does this level--or "threshold"--of category number vary with structure?
- How severe are the difficulties indicated?

Having examined the practical motivation for studying complex conditionals, let us turn to motivations from theory. There is almost no information in the psycholinguistic literature about the simple question of numerical limits on processing

---

*In the typical sentence verification task, subjects are timed as they evaluate the truth of a given sentence and respond "true" or "false." The task we have adopted tests verification time indirectly through responses to commands which are contingent on the antecedent truth value.

sentence categories.* This lack suggests that there are other motivations for the study we have conducted. Indeed, a review of the data base for theories of sentence comprehension, searching for gaps in current research, may lead us to the same questions as the practical ones posed here.

The following section summarizes the relevant theoretical literature, showing what is already known about the questions of concern and revealing the knowledge gaps which our study addresses.

Theoretical Motivation for the Study: The Limits of Previous Research

One of our goals in reviewing the theoretical literature below is to cover the research extending sentence comprehension from simple to coordinate structures of increasing complexity. This ground is largely unexplored, as our review will show. Because comprehension has been most thoroughly examined in the verification paradigm, and because verification is the core task in the instructions we are concerned with, we begin our review with research in that paradigm. Not all the variables of interest have been examined in the verification paradigm. So we further cover the diverse evidence on the effects of these variables in other paradigms. Additionally, we review what is known about certain of our dependent variables: particularly, how subjects create strategies to help process difficult sentences.

In preview, the lines of research we will survey and the questions we address are these:

1. Sentence comprehension in the verification paradigm. What is known about how comprehension and verification proceed in sentences with more than one underlying proposition? How

*The seminal 7-plus-or-minus-2 proposal of Miller (1956) is too general to allow precise prediction in this situation.

far does this research go? What sentence variations and logical functions have been examined, and what is their effect on comprehension?

2. Logical rules in the concept-learning paradigm. What is known about how logical concepts are grasped--particularly or and and--and their order of difficulty? Does any other work support this ordering?

3. Reasoning processes: Evidence and theory. What is known--and in relation to what sentence types--about how subjects handle extended logical operations, such as truth table calculations, that exceed what is usually considered necessary to understand a sentence? Have coordinate sentences, those using and and or, been examined?

4. Negation. What is the effect of negation on sentence processing generally? What is known about this effect in longer sentences with more than one negative term?

5. The nature of sentence reference in the verification paradigm. Has personal ("you are . . .") reference been investigated in sentence verification? With what results?

6. Partial processing in sentence comprehension strategies. Under what conditions do subjects adopt strategies that shortcut sentence processing?

Sentence comprehension in the verification paradigm. Understanding a novel sentence can be characterized as a compositional process (e.g., the "semantic composition" of Rips, Smith, & Shoben, 1978). The process involves accessing the meanings of the constituent words and combining them to yield a composite meaning for the sentence. This combining activity ordinarily consists of extracting the syntactic and logical relationships among the words of the sentence and using them as a framework for constructing sentence meaning.* An intervening step in semantic

*Frequent exceptions to this process of comprehension have been observed. For some tasks, comprehension may proceed by semantic strategies that by-pass syntactic structure (Fodor, Bever & Garrett, 1974; Bever, 1970). Other strategies consist of partially processing the sentence string to meet particular task demands.

composition for complex (multi-proposition) sentences is building the subordinate semantic units--the underlying propositions--that compose the larger sentence.

The form which this composite meaning takes has been hypothesized in theories of memory and comprehension. Kintsch (1974) and Anderson (1976) have modeled internal representations of sentence meaning in their work on semantic memory. Clark (1973), Clark and Chase (1972), and Carpenter and Just (1975) have depicted internal sentence representations in theories of the mental processes underlying tasks such as verification. Though differing in several respects, these theories posit in common a semantic representation which encodes the sentence in abstract propositional form.* Guiding this representation is a "deep structure assumption" that each idea in the sentence is a separate underlying proposition. The structural relations (subordination, coordination) among the component propositions are preserved in the final sentence representation.

The example below demonstrates the formal aspects of internal semantic representation for the sentence, "Canaries have yellow wings." The final format is Kintsch's (1974).

> Sentence -- Canaries have yellow wings.
>
> Propositional breakdown -- subject + object: Canaries have wings (have, canaries, wings); object + adjective: Wings are yellow (yellow, wings)**
>
> Final representation -- structurally related propositions --(have, canaries, wings) & (yellow, wings)

---

*A proposition is a unit of meaning that contains the functional relations present in a simple sentence: subject-predicate, etc.

**Kintsch's representation places the verb or attribution relation before the sentence arguments.

The order of the propositions in the final representation indicates syntactic subordination of the second proposition, as in a phrase-structural description. The "&" relation indicates logical conjunction (or intersection) which implies that both propositions must be true for the sentence to be true.

Using the sentence representation in tasks such as question-answering and deciding truth or falsity again brings the composition process into play, according to the conception of Rips et al. (1978). To confirm a two-proposition sentence such as (1) "Canaries have wings and beaks," we first evaluate the truth of the component propositions ("Canaries have wings" and "Canaries have beaks"), as well as the relationship implied by "and," and from those results derive the composite truth value for the sentence. Assuming that sentence verification proceeds by just such a combination of propositions, Rips et al. focused on sentence frames of the garden variety subject-verb-adjective-object--such as (2) "Canaries have yellow wings"--to investigate the order in which the component evaluation procedures apply. Reaction times confirmed that the verification process used by subjects does analyze and verify the sentence propositionally.* Results further suggested that subjects' verification strategies proceed in a semi-predictable order of propositions, reflecting partial processing by terminating when a false proposition occurs.**

---

*In fact, averaged over subjects, the verification process was found to recognize three propositional components for such sentences. The third proposition, in addition to the two traditionally recognized by transformational grammar and by an analysis like Kintsch's is: subject + adjective ("Canaries have yellow (X)"). Subjects' verification strategies consistently selected two of the three propositions for priority in processing: subject + adjective and subject + Object (counter to predictions assuming syntactically dependent processing).

**The order of processing propositions is only semi-predictable because of a pervasive strategy-shifting observed among subjects.

In probing sentences that relate three categories (canaries, wings, yellow), Rips et al.(1978) extended the investigations of simpler sentence verification (two-category, one-proposition, such as "Canaries have wings") which have dominated research on semantic memory (e.g., Collins & Quillian, 1970; Loftus, 1973; Smith, Shoben & Rips, 1974). Further, the Rips et al. query focused on the processes involved in comprehension and verification rather than on the structure/distance relations in conceptual memory addressed by the Collins and Quillian tradition. The Rips et al. study, however, along with the entire earlier line of investigation, constrains what we can learn about sentence processing by limiting the central variables to lexical relations within a single syntactic frame.* What happens when we are faced with longer and more diverse sentences -- those which relate a greater number of categories and in which the linking relationships take the form of logical rules of various types?** For example, what difference does it make for the verification process when an "or" links the underlying propositions rather than the "and" expressed in concrete and abstract form (respectively) in sentence types (1) and (2) above? One difference is the obvious ineffectiveness for sentences with disjunctive propositions of the self-terminating strategy which Rips et al. observed for conjunctive relations.

There is a second line of sentence verification research that brings us closer to answering the question of how we process sentences in which the conceptual relationships are defined by logical rules of varying types and complexity.

---

*Lexical variations include comparisons such as "yellow wings" vs. "yellow legs."

**We assume that syntactic rules are the essential functional relations that define sentence structure: subject-predicate, verb-object, etc.; logical rules define conjunction, disjunction, implication, and other deductive concepts which are not essential to sentence structure. We should also distinguish sentence comprehension from reasoning processes, though the distinction is not well-defined.

Research into the mental operations of sentence comprehension has used the verification paradigm to investigate the effect on reaction time of variations in the sentence frame itself. Begun by Wason (1961), extended and formalized by Trabasso, Rollins, and Shaughnessy (1971), Clark and Chase (1972), and Carpenter and Just (1975), this work has tested syntactic variations such as the active-passive and the cleft transformations; lexical variations such as markedness; and logical-semantic variations such as negation, comparison (two- and three-term), and quantification (universal and particular).

All these factors have been found to affect the difficulty of verifying a sentence. Accounting for this difficulty are the ad hoc comprehension models variously proposed (e.g., Clark & Chase, 1972), which formalize sentence verification as discrete operations occurring in three to four serially ordered stages. These models predict the time required for each stage and thus for the final evaluation response. The initial stage, termed sentence encoding, transforms the stimulus sentence into abstract propositional form. The time required in this stage is affected by each of the variations above. For example, a negative element in a sentence is encoded as the embedding of a core affirmative proposition within a negative one: "A canary is not a plant" becomes (It is false (canary is a plant)); whereas the affirmative version of the sentence is encoded more simply, as a single proposition. It is such variations in propositional complexity which are assumed to influence the subsequent stage of verification: comparing the sentence representation with information about the referent (e.g., canaries), represented in analogous propositional format.* The comparison stage is conceived of as a compositional process:

*The representation of the referent may be drawn from semantic memory, as in the canary example, or from an experimentally given picture, as when the sentence describes a specific situation ("The circle is above the square").

the propositions previously encoded guide a series of matching operations, which compare components between sentence and referent representations and check for identity or discrepancy. The comparison operations contribute independently to the time needed to compute a composite truth value.

Of the sentence frame variations employed in the verification paradigm, it is negation and quantification which complicate the sentence logically. To review the model, this complexity is depicted by representations which isolate the logical operators as superordinate propositions and embed the remaining proposition(s).* This representation in turn affects the complexity of the matching operations required to judge sentence truth or falsity. The verification models assume that the complex representations produced by negation, quantification, or other sentence modifiers, will always increase verification time, since these structures prohibit partial processing, in contrast with coordinate structure. The sentence type examined by Rips et al. (1978), subject-verb-adjective-object, permitted subjects to take processing short cuts by way of the underlying coordinate link between propositions.

The research supporting sentence verification models is limited, first, in looking generally at one sentence variation at a time (e.g., affirmation-negation); second, in restricting negation to a single occurrence;** and finally, in relating no more than two categories in a sentence. Thus the typical experimental sentence has not surpassed one or two underlying propositions in complexity. To look more closely at the complexity of the logical relation itself, and how it affects

---

*Particular and universal quantification, examined by Just (1974), are encoded similarly to negation: "Some canaries are plants" becomes "(Some canaries [canaries are plants])."

**An exception is Johnson-Laird's (1970) inconclusive study on the interpretation of doubly-negated sentences containing quantifiers.

verification, we must consider a wider range of logical rules applied to a greater number of categories within a sentence.

Logical rules in the concept-learning paradigm. The relative complexity of logical rules definable on two categories has been the topic of one branch of concept-learning research.* Early research on the experimental acquisition of bi-dimensional conceptual rules, using specially constructed stimulus referents, revealed a reliable order of rule difficulty (from easy to hard): simple affirmative ("the item is square"), conjunctive ("the item is square and red"), disjunctive ("the item is square or red"), conditional, and biconditional (see Neisser & Weene, 1962; Haygood & Bourne, 1965; Bourne, 1967). The conjunctive-disjunctive ("and - or") difference persists in the wider research on classification processes. Subjects show consistent biases toward selecting, and greater facility in learning, conjunctive over disjunctive classifications (e.g., Bruner, Goodnow, & Austin, 1956; Hunt and Hovland, 1960; Bourne, Ekstrand, & Montgomery, 1969). Supporting the rule-learning results, the language acquisition literature indicates that "and" precedes "or" developmentally (see Braine, 1976). Further, Rose and Cox (1980) offer an intuitively appealing explanation of the general cognitive advantage of "and," by analogy to electronic circuitry.

The work on rule-learning is unconcerned with the comprehension and use of sentences expressing conceptual rules. The typical experimental task presents the subject with a succession of stimulus items which are instances of the rule, and requires the subject to demonstrate concept attainment by consistently identifying further instances. Our concern, on the other hand, is not with acquisition but with the interpretation and application of rules presented verbally. Though we would expect

*Concepts are defined in this tradition as the relationship between categories (stimulus attributes or the combination of attributes) as characterized by logical connectives--e.g., "X or Y" expresses a disjunctive concept.

the order of difficulty found in acquisition tasks to generalize to application tasks, the "and-or" distinction remains to be examined in sentence comprehension theory and research.

A second gap in the research on logical rules occurs around extending "and - or" relationships from two to varying numbers of categories.

Reasoning processes: Evidence and theory. Processing sentences in experimental tasks such as verification can entail a complex chain of reasoning, exceeding the operations normally seen as sentence comprehension. The normal product of comprehension is a semantic description derived by extracting from the sentence the structural (propositional) base filled in by lexical material and including semantic relations, such as agent and object. This is the outcome of sentence encoding as conceived in the psychological models of Clark, Trabasso, Kintsch, and others, discussed earlier, and linguistic models such as that of Katz (1972) or that of Fillmore (1968) (though the latter differ in the form posited for semantic descriptions). A more extended processing appears to be the case when the sentence expresses a complex relationship among categories -- as reflected, for example, in determining what combination of categories or their negations makes the sentence true.

A well-studied example of extended reasoning is the conditional relationship expressed in "if-then" sentences, presented to subjects in various sorts of problem-solving tasks such as syllogisms. In order to solve the problem, subjects must go beyond understanding the given sentence--which is simple enough with the "if-then" construction in natural language--and deduce from that sentence the set of true and false states (combinations of propositions) defined by the truth table of material implication.* This deduction is notoriously difficult

*A crucial distinction must be noted here. The sentence type we are considering -- the conditional imperative -- is not an expression of material implication, and its interpretation is not subject to the same problems as the latter. The difficulty in the conditional imperative is interpreting the relationship of categories within the antecedent.

to make (Johnson-Laird & Tagart, 1969; Wason & Johnson-Laird, 1972; Johnson-Laird, 1975).

Sentences with three-term or larger linear comparisons have also been studied.* These sentences require extended reasoning when they are used in certain kinds of question-answering tasks (Clark 1969, 1972; Griggs & Shea, 1977). Finally, there has been some work on quantifiers, as mentioned in our review of verification. The inordinate difficulty of using sentences with two quantifiers ("Some X's have few Y's") in a range of verification contexts (Johnson-Laird, 1970a) suggests that these sentences may require logical calculations beyond extracting the basic sentence meaning.

To reinforce the distinction we are drawing between comprehension and reasoning, we should note that varying the linguistic structure appears to have little effect on the logical complexity of statements of relationships such as quantification or material implication. Attempts to reformulate the language of the conditional or of doubly quantified sentences have failed to improve substantially the reasoning errors that pervade deduction tasks (see, e.g., Legrenzi, 1970; Johnson-Laird, 1970b).**

---

*Example: "John is taller than Bill; Jack is taller than John."

**Comprehension is not a unitary phenomenon. The boundary between meaning and use, and meaning and logic is one of the more disputed questions of linguistic and psycholinguistic theory. In psychological theory, the question of what to include in a semantic description--i.e., what depth of processing to infer--has recently been resolved by deferring to the question "comprehension for what?" Thus the reader-hearer's purpose determines the form in which sentence meaning will be represented. In linguistic theory, the sentence base--the input to the semantic component--has been extended from grammatical relations (Chomsky) to conceptual relations (the generative semanticists and Fillmore) to abstractly related inferences and other extra-sentence propositions (Lakoff). The question of when to stop remains unsettled. We are not advocating any one viewpoint and leave open the question of what generic processes are involved in mapping a complex sentence onto a referential domain-- whether comprehension (meaning) or reasoning (logical representation).

We have chosen to focus on reasoning processes engaged by rules of a medium degree of complexity--as expressed by coordinate conjunctions, "or" and "and," For example:

(3) The item is either red or square, and either hollow or not smooth.

Performance with logically less and logically more complex structures (e.g., simple sentences on the one hand and conditional relations on the other) has been experimentally explored, as we have shown, but this middle ground has not been probed. The operations of conjunction and disjunction, when combined with negation in multi-category sentences, should produce the same sort of reasoning burden as conditional relations, though to a lesser degree. Verifying the sentence should be an intricate process of deriving the possible category combinations defined by the rule, then matching those possibilities against similarly encoded reference information to determine overall truth or falsity. This process goes beyond the steps outlined in traditional sentence comprehension models to encompass reasoning operations such as constructing a mental truth table.

We are assuming these logical operations will affect both the encoding and comparison stages of verification. For example, the logical meaning of "or" must ideally be captured at the stage of sentence encoding to enable accurate application of the sentence in a verification task. This means constructing the set of compound propositions (the distinct combinations of true and false categories) which make the sentence true according to the rule definition. For the relation integrating "or" and "and" in sentence (3), a set of nine combinations of individual category truth values (representing the presence or absence of the four item attributes) characterizes the truth conditions for the sentence (assuming inclusive interpretation of "or"). These conditions can be displayed in a truth table like the following, which evaluates the various combinations

of true (+) and false (-) categories.* This table displays just the four combinations which make the sentence true.

Table 1

Combinations of Individual Category Truth Values That Make Sentence (3) True(a)

Category	"True" Combinations			
red	+	+	-	+
square	+	-	+	+
hollow	+	+	+	+
smooth	-	-	-	+

(a) The true or false categories here can be considered to represent attributes which are present or absent in a referent object.

In deciding the truth of a sentence like (3), it may be that subjects use a mental schema something like the table above to generate the set of valid category combinations (represented as conjoined propositions), which would subsequently be compared against a representation of the referent (in this case, an experimental stimulus). A match between the referent and a member of the truth set would yield a "true" decision. Again, truth tables are not typically considered components of a semantic representation. Nor are operations on truth tables included in current models of sentence comprehension. There is simply little evidence or theory against which we can compare our conception of processing multiple coordinates.

Braine (1978) offers a specification of what he terms "inference rule schemata," intended to characterize operations in natural logic -- such as reasoning with "or," reasoning

*This is not truly a truth table, but a mapping of attributes onto the concept defined by the rule "A or B, and C or not D." The mapping table can be readily converted to a truth table, which relates propositions. (Each category is a proposition in deep structure.)

with "and," eliminating alternatives, detecting contradictions, and deducing the truth sets of material implication. These formulae are of help in defining the readings of "or" and "and" and in showing the conclusions to be reached from given premises. But Braine's specifications do not suggest how the schemata are combined and used, which is our chief concern here.*

Negation. Previous research convinces us that negative sentences ("It is not A.") are harder than affirmatives ("It is A.") in most instances (see summaries of the evidence in Wason, 1962, and Trabasso, 1972). Although the difficulty disappears in certain contexts (e.g., plausible denial, Wason, 1965; change in meaning, Greene, 1970; and other situations that are natural to negation, Wales & Grieve, 1969), the general conclusion holds for the majority of experimental tasks. For example, negative sentences are harder to recall (Miller, 1962; Mehler, 1963; Clark & Card, 1969) and to match (Miller & McKean, 1964); they complicate three-term problem-solving (Clark, 1969); and they take longer to verify against pictures of semantic memory (Wason, 1959, 1961; Wason & Jones, 1963; Gough, 1965, 1966; Slobin, 1966; Trabasso et al., 1971; Clark & Chase, 1972; Chase & Clark, 1972; Anderson & Reder, 1974).

The negation effect is robust. In implicit form--e.g., "doubt" for "think not"--negation appears to increase difficulty fully as much as its overt expression (Sherman, 1969; Just & Clark, 1973; Just & Carpenter, 1975). This result generalizes to implicit negation in quantifiers ("few" for "not many," Just & Carpenter, 1971).

Early explanations of the difficulty of negation pointed to transformational complexity (Miller, 1962; Miller & McKean,

*See also Johnson-Laird (1975) for a general discussion of natural reasoning.

1964; Mehler, 1963; Gough, 1965); while later explanations (Clark & Chase, 1972; Trabasso et al., 1971; Carpenter & Just, 1975; with preliminary work in Wason, 1962) combined linguistic with psychological factors, positing the mental processes by which negative sentences are encoded and used in particular tasks. By the processing explanation, verifying a negative expression requires an extra operation in matching sentence to referent.

How negation varies with other factors has been studied to a limited extent. The scope of negation (the size of the unit modified) has been found to affect the complexity of negative sentences (Carpenter & Just, 1975), as has the sentence truth value. The latter effect is well replicated (see citations in the Carpenter & Just review, 1975): specifically, false negatives are usually easier than true -- the reverse of the order for affirmative sentences.* Position of the negative word in the sentence has also been varied, but with equivocal results (Jones, 1968; Johnson-Laird, 1970a; Carpenter & Just, 1975).

The number of negative words in a sentence has been looked at indirectly, and the results are tentative. Not surprisingly, adding negations appears to create successive comprehension decrements (Sherman, 1976; Charrow & Charrow, 1978, 1979; Johnson-Laird, 1970a; Legrenzi, 1970). But critical to the number of negatives in a multiply negated sentence is the structural relationship between those words: whether they relate independently (e.g., "not X and not Y") or hierarchically (one within the scope of another, as "not both X and not Y"). Johnson-Laird, using double negation of the hierarchical type ("No X are not Y"), found that such structures severely impair comprehension unless subjects adopt a translation strategy, recoding the two negatives as affirmatives.

*But see Trabasso et al. (1971) for evidence that the true-false effect depends on whether sentence or picture is presented first in the verification task.

But the amount and structure of negation have not been systematically manipulated. Nor has the possible interaction been tested between negation and sentence length and complexity. Finally, the effect of these variables on errors in sentence verification and other comprehension tasks has not been probed, the traditional negation studies having tested primarily reaction time. It is with these unexplored areas that we are concerned here.

The nature of sentence reference in the verification paradigm. A final aspect of verifying logically complex sentences that is relevant here is the nature of the information against which the sentences are to be evaluated. The verification tasks we have cited, from that of Wason (1961) to that of Rips et al. (1978), typically refer the stimulus sentences to one of two kinds of items: information stored in semantic memory (e.g., Rips et al., in the tradition of Collins and Quillian) and specially constructed experimental stimuli. The latter take the form of abstract figures (e.g., Clark & Chase, 1972) or pictures of events (e.g., Gough, 1965, 1966).

Our interest is with a third type of reference--information stored in memory about oneself. Personal meaning of this sort, which varies with the particular referent, is one form of what linguists call deixis, or location-specific reference. The particular deictic information we are concerned with in this study resembles general semantic memory--basic, overlearned, frequently used categories (e.g., age and sex)--more than episodic memory--specific events with time and place indices.

In keeping with the verification models described earlier, we assume that whatever the nature of the referent, it is internally represented for the verification task in a unified format: an indexing of propositions that permits direct comparison with the sentence representation.

Partial processing in sentence comprehension strategies. The types of sentences we have chosen to look at are coordinate structures, and we are further assuming that comprehension/verification of the sentences is a compositional process -- thus constituting conditions sufficient for partial processing. For example, a sentence expressing multiple alternation, once scanned to detect the presence of "ors," can be processed to the point of obtaining a single true proposition and then terminated. Or a sentence with multiple iteration ("and") need be processed no further than finding a false proposition, as was the case for the Rips et al. subjects. Because sentences with the complex combinatorial logic which is our focus have not been studied, we have little prior evidence from which to formulate hypotheses about processing strategies. We can only indicate the logical possibility for partial processing and point out evidence for the use of that strategy with other kinds of tasks and stimuli. Whether those processing short cuts which are structurally possible will actually be adopted depends on a variety of primarily non-linguistic factors which have not been widely studied.

Recapitulation

We have covered the theoretical background for our inquiry into instructions with logically complex conditions, as well as the relevant prior research. Quite simply, there is very little evidence for the questions we proposed earlier:

- What factors--at what levels and in what combinations--contribute to the difficulty of multi-category conditional sentences?
- How great are these difficulties, and in particular, do there exist thresholds of sentence complexity beyond which comprehension deteriorates markedly?

In response to these questions, our study has attempted (1) to diagnose the difficulties and classify their sources,

(2) to bound performance problems in terms of the source variables, and (3) to explain the problems in terms of subjects' processing strategies vis a vis intrinsic logical complexity.

Description of Stimuli and Specific Hypotheses

To investigate comprehension of multi-category sentences, we used a modified sentence verification task, measuring the time and errors involved in performing an instruction. The instruction depended on assessing the truth of an antecedent statement. We have already outlined general characteristics of the stimulus sentences selected for this task. Recall that these sentences are of the form "If you are X, do Y," where "X" is the coordinate compound of several categories.* It is this compound which contains the variables of interest: manipulations of logical and syntactic structure. The categories themselves are familiar, one-word concepts (e.g., "male," "employed") selected for ease of processing and minimal interference with structural effects. Below, we will describe the structural variables and how we selected them.

We began with a classification of compound statements designed to reveal differences in difficulty and the sources of those differences. First, we looked for the dimensions that define logical meaning (or structure) in this domain. Specifically, the domain of interest encompasses all statements with compound categories joined in coordinate relationships--that is, categories linked by "and" and "or" rather than by subordinate relations such as "because" and "when." Standard logic (e.g., Quine, 1959) calls such compounds "truth functions" because the truth value of any one of them can be determined

*Each category represents a proposition in mathematical logic as well as in linguistic theory.

purely from the truth values of its component categories, regardless of the meaning of those categories.*

We observed four structural dimensions in this domain, which when combined describe the set of all logically distinct meanings. Three elementary operations are recognized by standard logic in making truth functional compounds: negation, conjunction, and alternation.** These operations correspond to "not," "and," and "or" in natural language. "And" and "or" form a complementary pair: they are the alternate ways of connecting one coordinate category with another. This analysis gives us two basic dimensions: affirmative-negative and "and-or." Standard logic recognizes another determinant of logical meaning: grouping of categories. Thus, "(A and B) or C" may differ in truth value from "A and (B or C)" (with grouping indicated by parentheses). Alternatively, some structures -- e.g., "A and B and C" -- are ungrouped. A final essential variable is the number of categories connected in a compound.

We arrived, then, at four conceptually independent factors to classify coordinate structures. Three of these factors concern relationship; one is a matter of counting. In summary, these factors are:

1. Number of categories contained in the antecedent statement;

---

*The coordinate-subordinate distinction is grammatical. In logic, subordinate relations are not considered "logical connectives": they don't affect truth value. Nor do shades of meaning among coordinate conjunctions, such as "but" and "as well as," count as logically distinct. It is just those distinctions affecting truth value that interest us here.

**These operations, or "connectives," are ways to form statements from other statements. Other connectives include "if... then" and "unless," derivable from the basic operations. We note that these operations compose the language of propositional logic, as distinct from predicate logic. The latter consists of operations such as the quantifiers "every" and "some,"and defines the internal structure of propositions rather than how they are connected.

2. Organization of categories: whether or not they are grouped, and if so, how;

3. Nature of connective between categories: whether it is "and" or "or";

4. Presence or absence of negation operating on a given category.

We defined a fifth factor, nested under organization (Factor 2). This specified, for grouped structures, the manner of grouping (e.g., whether there were two or three clusters in a given sequence of categories), and for ungrouped structures, how the connective was expressed (whether explicitly repeated or implied with commas). Unlike the others, this last variable is purely a distinction of surface grammar.

Although we have spoken of the main factors as logical, each can be described linguistically as well. Grouping is analyzed as hierarchical branching in phrase structure grammar. (Phrase structure does not, however, isolate logical distinctions.) Negation is a syntactic property, as is number of constituents. "And-or" can be considered a semantic as well as a logical distinction, and the more general coordinate-subordinate distinction of grammar reflects the truth functional vs. non-truth functional compounds of logic.

The typology of variables we have defined excludes some dimensions with potentially large effects on sentence difficulty in this domain. We focused on basic logical types while ignoring various ways of wording those types--that is, variables such as active-passive, reordering of coordinates, or logical equivalents (e.g., "A, and B or C" vs. "A and B, or A and C"). The practical question guiding our selection in this study was: given the set of distinct logical structures, which types are so complex as to suggest a need for alternative forms, such as rewording or non-prose presentation?

We have named the factors necessary to distinguish coordinate structures. Our stimulus set selectively samples these

structures. To define this set, we chose those levels of the designated factors occurring most frequently in real documents, or those assumed most likely to affect processing. These assumptions we based first on indications from prior research and second on the results of a pilot study which suggested some specific sources of difficulty in compound conditionals.

We will describe in detail the five variables we manipulated and the levels defined for each, including the rationale for selecting specific levels and the associated predictions about difficulty.

At the end of this section, we will present a table which shows the total set of structures resulting from the factorial combination of each of the five experimental factors. (Because the description of these factors is complex, the reader may wish to refer to this summary, Table 3 on page 37, throughout.)

Factor 1: Category Number. This factor had four levels, corresponding to whether four, five, six, or seven categories comprised the antecedent.

Predictions. As both intuition and past research suggest, increasing the number of decisions to be made in decoding a sentence (or any other unit of processing) should decrease speed and accuracy of processing. However, we have no evidence and no predictions for how many decisions subjects can handle and what precise effects multiple decisions would have on speed and accuracy.

Rationale for defining the levels. Concerning error data, we observed from the pilot results that the hardest sentences seemed to produce errors beginning at four categories while the easiest were accurate until around six. Four to seven categories thus seemed a useful range to explore.

Factor 2: Organizational Type. This factor had three levels, specifying whether and how categories were grouped. We illustrate each level with three-category sentences.

- String ("and" or "or"). This structure is ungrouped. All categories are connected at the same level and by a single type of connective. For example: "If you are male, married, and employed, do X."

- Simple grouping ("and-or" mixed). This structure divides the string into groups of categories, creating hierarchical branching (in phrase structure grammar terms). Simple grouping is manifested by a mixture of connectives: "If you are male, or married and employed, do X."*

- "Not" grouping ("either-both"). This structure also groups categories but on a logically different basis from the simple hierarchy. It results from negating a two-category compound: "not both X and Y" or "not either X or Y." (In logic these structures are called alternative denial and joint denial, respectively.) A hierarchy created in this way does not require mixing connectives: "If you are male and not both married and employed, do X." (We distinguish "not" grouping from the negation variable: The latter negates a single category; the former, a compound. Simple negation is independent of and can combine with compound negation: "not both single and unemployed.")

Predictions. We expected the following order of difficulty from easy to hard: string, simple grouping, "not" grouping. Because these organizational types are nominal rather than ordered, our prediction derives from comparing them in pairs.

First, groupings should be more complex than strings because (1) the hierarchy places a burden on short-term memory during left-to-right sentence processing (See Yngve, 1962), and (2) the change in conjunction activates a new logical operation (a "processing subroutine;" Smith, 1970) different from that used to interpret the primary relationship. The second of these rationales is straightforward. The first, based on the work of Yngve and of Johnson (1965, 1968), asserts the reality of surface structure and requires illustration: A reader who sees a clause beginning "It is A and either..." expects at least two more categories to occur and, according to Yngve, must hold the expectation in memory while interpreting the beginning fragment of the clause. "It is A and B..." entails no such syntactic commitments.* More generally, hierarchies carry a greater memory load because they contain more levels: Processing a string requires keeping a single level of relationship in working memory; a grouping requires remembering two or more simultaneous levels.

As the second part of our prediction, "not" groupings should be harder than simple groupings because of the logical requirements of joint and alternative denial. Haygood and Bourne (1965), and the line of research stimulated by them, found the latter rules harder than simple conjunction in concept-learning and attribute identification tasks. A single study (Rose & Cox, 1980) has examined two-category compounds in a verification task and found that negated compounds take around twice as long as the corresponding affirmative compounds. That the "not" grouping involves negation is another reason for predicting its pre-eminent difficulty.

---

*Despite the fact that the particulars of Yngve's model are questionable (Rohrman, 1968; Perfetti, 1969), the reality of phrase structure cues in sentence processing, established by Yngve's and by Johnson's work, has held up during a decade of research.

Rationale for defining organizational types. In considering antecedents with coordinate categories, we are faced with many ways to classify them structurally. The simplest first cut to make looks at whether or not the coordinates form subsets (i.e., different levels of relationship). This variable is a good candidate for affecting difficulty, since complexity can reasonably be defined by the number of levels to be processed. This dimension gives us the string (one-level) vs. grouping (two-levels) distinction.

The third organizational type, "not" grouping, lies on a different dimension. We justify classifying it separately from the simple grouping on several grounds. First, it entails a separate set of sentence subtypes, as specified by the fifth factor in this study. Also, "not" hierarchies consist of more than merely simple hierarchies with negation added to the groupings.*

Factor 3: "And" vs. "Or." This factor had two levels, specifying whether "and" or "or" was the main connective in the compound. The main connective for hierarchies is the one linking the highest level of category groupings. The sentences below show the main connective by underlining:

(a) You are male, and employed or married.
(b) You are male and employed, or married.

Sentence (a) represents the "and" dimension; (b) represents the "or" dimension. A string simply repeats the main connective throughout.

Predictions and rationale. "And should be easier than "or" in all cases. The presentations of conjunctions and alternation

*First, the required mixture of connectives is different between the two types of hierarchies ("not" hierarchies do not require a change in conjunction). Second, the compound "not" within "not" hierarchies is restricted to groups of two categories, whereas simple hierarchies allow wide variation in size of groups. (Examples under Factor 5, p. 32, will clarify these two distinctions.)

in standard logic (see Quine, 1959) provide no a priori ordering of the two operations as to complexity. Nor do traditional linguistic representations. We base our prediction on the empirical ordering obtained in the somewhat scattered lines of research reviewed earlier. Additionally, a recent study by Rose and Cox (1980), using a modified sentence verification task with two-category conditional instructions, found sentences with "or" took consistently longer than sentences with "and." While the "and" advantage has yet to be established in sentence comprehension research or theory, we have cited some general reasons to expect it to emerge in this study.

The prediction is clearer for strings than for grouped structures. The latter combine "and" and "or" within the same sentence, thus contaminating the distinction, as shown in the examples above. It was the pilot results that suggested the "and-or" asymmetry would transfer to hierarchies and would emerge through main rather than subordinate connectives. In summary, we expected to find the "and" advantage universally, though with smaller effects in structures where both connectives occur.

Factor 4: Affirmative-Negative. This factor specified whether any categories were singly negated--e.g., "not married and not employed." We will characterize such negation as simple to distinguish it from the compound negation defined earlier ("not both married and male"). The presence of simple negation was defined as two negated categories in the antecedent, with "not" as the negative word:

You are male, not employed, and not married.

We placed two restrictions on the occurrence of simple negation in the stimulus sentences. Because of the ambiguity arising from negating a non-terminal category (e.g., "You are not male, employed, and single" seems to negate the triple)--an ambiguity clarified only by adding potentially confounding words or punctuation--we restricted the location of simple

negation to the last two categories in the antecedent (as in the sentence above). Also, because of the inordinate difficulty and apparent anomaly of cyclical "not," as in "not both male and not married," we prohibited simple negation from occurring within compound negation.

Predictions. Sentences with negated categories should be harder than those without. The grounds for this prediction are well-known and are covered in the negation section of the review above.

Rationale for selection. We chose to define negation on two categories based on the pilot results and the findings of Rose and Cox (1980). These suggested that simple negation does not produce errors unless it occurs at least twice in a sentence. This result precisely fits our present concern: the reflection in errors, or in exceptional latency increments, of processing limitations given increases in complexity. There was a practical reason for this choice as well. Our attempt to diagnose difficult sentences in real-world documents turned up two (or more) negations as a common characteristic. Yet the typical specification of negation in sentence comprehension research has been a single occurrence. (Recall that sentence comprehension research has focused on the simple sentence and neglected the compound, where double negation is more likely to occur.)

Factor 5 (Nested):Organizational Subtypes. The three organizational types defined under Factor 2 require further structural specification. Hypothesizing that some of these specifications would critically affect complexity, we manipulated a fifth variable, nested under type of organization, to sample two of the possibilities for each type:

(a) Simple grouping subsumes numerous possible arrangements, characterized on three dimensions: (i) the number and size of constituent groups, or main branches, (ii) the number of levels at which categories are grouped (levels of branching), (iii) the sequencing of those groups that differ in size or level. (This last dimension distinguishes, among other things,

right and left branching.) For a hierarchical structure of n categories, the number of possible groups on a given level ranges from 2 to n-1; and the size of groups (number of categories included), from 1 to n-1 (size and number are mutually determining). Other formulae enumerate the possibilities for the last two dimensions. To illustrate, we present a comparison on each dimension using a four-category compound:

(i) Number and size of groups

A, B, and C or D vs. A and B, or C and D

(3 branches: 1 + 1 + 2) (2 branches = 2 + 2)

(ii) Level

(A) and (B) and (C or D) vs. (A and [B or C] or D

(2 levels) (3 levels)

(iii) Sequence (by size and number)

1 + 1 + 2 vs. 2 + 1 + 1

We designated number of groups as the nested factor for simple groupings, introducing two vs. three branches into the structures in this class (illustrated in (i) above). (Keeping a fixed number of groups meant that size of group varied across total category number.) The other two factors, level and sequence, we held constant: the former since multi-level branching is unusual in the practical domain we are considering; the latter, since group order does not (for coordinate groups) constitute a distinction of logical meaning. Level of groupings was fixed at two, and groups were sequenced in order of ascending size (illustrated by the first structures in comparisons (i), (ii), and (iii), above).*

The following table summarizes, for each number of categories, how the number-of-branches variable structured the size and sequence of groups in simple hierarchies.

*The choice of ascending size respects the document design principle of presenting the most general distinctions first.

Table 2

Group Size and Sequence in Simple Groupings with 2 and 3 Branches

Number of categories	2 branches	3 branches
4	2 + 2	1 + 1 + 2
5	2 + 3	1 + 2 + 2
6	3 + 3	2 + 2 + 2
7	3 + 4	2 + 2 + 3

Rationale and predictions. We expected that, given an equal number of categories, the number of branches subsuming those categories would affect difficulty. There is little research to indicate the direction or extent of such effects. Based on a complexity metric that simply counts constituents, we predicted that three groups would be harder than two.

(b) "Not" grouping, as we have defined it, has more limited possibilities. This organizational type requires at least one two-category group -- the "not either/or" and "not both/and" phrases. With that constraint, this structure can vary on the three dimensions described for simple groupings plus one other. This last variation has two orthogonal realizations (assuming the main connective is given): (i) the nature of the compound negation (joint vs. alternative denial -- "not both" vs. "not either"); (ii) the relation between the main connective and that inside the "not" compound (same vs. different). The second property requires illustration:

same: A, B, and not both C and D.
different: A, B, and not either C or D.

We selected this two-dimensional factor to specify variations in "not" groupings. Values on the other three dimensions were fixed as follows. All structures in this class contained three groups arranged in order of ascending size, illustrated above as well as in Table 3. The "not" compound occurred as the left-most two-category group -- i.e.,

the last phrase in four-category structures; the middle phrase in five, six, and seven-category structures (see Table 2).* These sentences can be seen summarized in Table 3.

Rationale and predictions. Both aspects of this manipulation were assumed to play potentially significant, but as yet unexamined, roles in the comprehension of coordinates. We predicted asymmetric difficulty for the "either-both" as well as the same-different comparison. Specifically, as proposed earlier, we hypothesized that a change in conjunction causes confusion. Thus "and not either..." should be harder than "and not both...". Further, having found isolated evidence that alternative denial ("not either") is harder than joint ("not both") in two-category antecedents (Rose & Cox, 1980), we predicted this result would generalize to the present study.

(c) String subtypes: Explicit vs. implicit. In partitioning strings, we departed from variables distinguishing logical meanings, since the string is itself logically basic. We chose to manipulate how the distribution of the connective was expressed, suggested by the pilot study as a critical determinent of complexity. In that study, the typical, comma-implied distribution of the connective produced unexpected errors in the "or" condition. We attributed these to an "and" bias in interpreting the commas separating coordinates: "A, B, C, or D" tends to be read "A and B and C, or D" by some subjects. The pseudo-ambiguity of this construction (it is not grammatically ambigious) can be resolved by making the distribution of the connective explicit:

A or B or C or D.

*These values were not set arbitrarily. The choice of three structural divisions was necessary to obtain a two-category group in each of the four- through seven-category antecedents. The rationale for ordering by ascending size was explained earlier. The left-most location chosen for the "not" compound should have the most immediate effect on comprehension.

Predictions. For subjects susceptible to an "and" interpretation of "or" strings, the explicit form should be less difficult than the implicit. Specifically, if the implicit form is entirely misinterpreted (only one reading occurs), then the superiority of the explicit should be reflected mainly in reduced errors (with a slight time reduction corresponding to the difference between hierarchy and string). On the other hand, if the implicit is confusing (two readings occur), then the explicit should reduce both errors and time.

For "and" strings, we expected no difference in accuracy between explicit and implicit forms, but we did expect some effect on response time. These sentences might suffer from the redundancy of explicit distribution -- which is atypical, somewhat unexpected, and slightly longer than commas.

Summary. We have described five variables--four main and one nested--with the following numbers of levels:

(1) number of categories--four levels
(2) organizational type--three levels
(3) and/or--two levels
(4) simple "not"--two levels
(5) organizational type--two levels (nested under Factor 2)

Crossing all factors yielded 96 structures. Omitting the simple "not" variable, Table 3 shows half of these structures, with letters representing categories. Except for the 16 superficially distinct explicit-implicit pairs, these structures reflect logical distinctions in coordinate compounds. As the table shows, we inserted semicolons in the simple and the "not" groupings. These are necessary in seven- and six-category two-groups and in seven-category three-groups to avoid ambiguity. This arises in any structure with commas internal to groupings -- such as "A, B, or C, and D, E, or F" -- clarified by "A, B, or C; and D, E, or F." To equalize any semicolon effect (e.g., it might facilitate comprehension by emphasizing groupings)

TABLE 3
The Stimulus Set [a, b]

Organizational Type, Sub-Type, and Connective	NUMBER OF CATEGORIES 4	5	6	7
STRING				
and — Implicit	A, B, C, and D	A, B, C, D, and E	A, B, C, D, E, and F	A, B, C, D, E, F, and G
and — Explicit	A and B and C and D	A and B and C and D and E	A and B and C and D and E and F	A and B and C and D and E and F and G
or — Implicit	A, B, C, or D	A, B, C, D, or E	A, B, C, D, E, or F	A, B, C, D, F, F, or G
or — Explicit	A or B or C or D	A or B or C or D or E	A or B or C or D or E or F	A or B or C or D or E or F or G
SIMPLE GROUPING				
and — 2 group	A or B; and C or D	A or B; and C, D, or E	A, B, or C; and D, E, or F	A, B, or C; and D, E, F, or G
and — 3 group	A; B; and C or D	A; and B or C; and D or E	A or B; and C or D; and E or F	A or B; and C or D; and E, F, or G
or — 2 group	A and B; or C and D	A and B; or C, D, and E	A, B, and C; or D, E, and F	A, B, and C; or D, E, F, and G
or — 3 group	A; B; or C and D	A; or B and C; or D and E	A and B; or C and D; or E and F	A and B; or C and D; or E, F, and G
"NOT" GROUPING				
and — Same	A; B; and not both C and D	A; and not both B and C; and D or E	A or B; and not both C and D; and E or F	A or B; and not both C and D; and E, F, or G
and — Different	A; B; and not either C or D	A; and not either B or C; and D or E	A or B; and not either C or D; and E or F	A or B; and not either C or D; and E, F, or G
or — Same	A; B; or not either C or D	A; or not either B or C; or D and E	A and B; or not either C or D; or E and F	A and B; or not either C or D; or E, F, and G
or — Different	A; B; or not both C and D	A; or not both B and C; or D and E	A and B; or not both C and D; or E and F	A and B; or not both C and D; or E, F, and G

a We have omitted the negation variable in this representation. The affirmative forms above were compared with corresponding negative forms, in which "not" was placed before each of the last two categories ("A, B, not C, and not D"). The exception is 4-category "not" groupings, in which the first two categories were negated.

b The letters A through G represent categories in the antecedent, functioning as predicate nouns or adjectives (e.g., "If you are male and a homeowner.....").

we placed semicolons identically at group boundaries for all numbers of categories.*

Method

Materials

The stimulus materials were 96 sentence frames of the form, "If you are X, press (button) Y," where X was replaced by each coordinate structure generated by the five-way factorial combination described above (see Table 3). These sentences functioned as conditional instructions, with the antecedent ("if" clause) expressing a reader-referenced statement on whose truth value the command ("do" clause) was contingent.

Description of categories. The fillers for each structure were selected from a set of 15 highly familiar, one-word (or in one case, two-word) categories. Ten of these categories were paired as complementary antonyms. The total set of 15 words, with antonym pairs given first, was: "male-female," "married-single," "white-black," "righthanded-lefthanded," "over 25-under 26," "employed," "homeowner," "student," "parent," "veteran." (The last four categories were preceded by the article "a" in the stimulus sentences.)

Rationale for choosing categories. To get these categories, we sampled from the demographic and personal characteristics used frequently in documents, choosing those which, in our opinion, required only overlearned evaluation responses from most readers. We selected those characteristics which could be expressed briefly (preferably by one word), and with words that are specific, concrete, and familiar.**

---

*We did not use other ways of clarifying structure, such as repeating the subject ("You are A or B and you are C"), to avoid introducing confounding factors in comparing sentence types.

**We developed no metric for specificity or concreteness but relied on our judgment of fairly obvious distinctions. "Old," for example, is brief and familiar but vague compared with "over 25."

By simplifying categories, we intended to minimize their processing relative to processing sentence structure. Finally, we chose half the concepts as antonym pairs to enable assigning any S a subset of 10 words (from the 15 total) with the truth value distribution described below.

Sentence generation. The sentence categories were filled according to the following procedure, repeated between subjects. Ten words were selected from the set of 15 such that half were true and half false when applied to the subject. Ordered sets of these 10 were then randomly drawn (without replacement) for each sentence frame in the set of 96, with the restriction that half the resulting antecedents were true and half false for that S. Repeating the randomization procedure between Ss prevented potential interactions of categories with structural type; however, it was not possible to equate the probability of any sentence type being true or false across subjects. Thus, each sentence type eventually had a different number of trues and falses when viewed across subjects.

The imperative clause ("press Y") of the experimental sentence frame instructed S to press one of two buttons, A or B. For each S, the two commands were equated across frames and counterbalanced with true and false trials.

These procedures produced 96 conditional instructions incorporating the 96 structural types. There were no repeated categories within sentences, and category selection was randomized across types and Ss.

Implementation

A computer program generated the 96 sentences in an interactive session with each subject. The program was run on an HP3000 system, to which a 2640B display terminal was connected. Besides completing individual sentence frames according to the procedures described above, the program was designed to order the completed sentences randomly with the following restriction:

sentences of the same length or same organizational type could occur no more than three times in succession. Each S received a different random ordering.

The program also scored and recoded response time and accuracy.

Procedure

Ss were tested individually in a single session lasting 35-45 minutes. Each S responded to all 96 instructions, presented one at a time on the screen of the HP display terminal. Subjects sat at the keyboard of the terminal, where they indicated their responses by pressing one of three buttons, labelled "A," "B," and "NA" (not applicable). The rest of the keyboard was covered. Ss were instructed to respond with the index finger of the preferred hand throughout the trials. The three buttons were about 1.5 inches equidistant from a center point marked "start," where S rested the response finger between trials.

Ss were read instructions explaining that this was a study to see what kinds of sentences are harder or easier to understand. They were asked to read and respond to each sentence by pressing the appropriately labelled button. This was "A" or "B," whichever the sentence instructed, if the conditions described were true for the subject; "NA" if false. Ss were instructed to read and respond as fast and as accurately as possible.

Before trials began, Ss saw 10 questions on the screen representing the 15 categories ("Are you male?" etc.) They were asked to answer "yes" or "no" while E typed in the answers. The program then selected 10 words (out of the 15 possible) yielding an equal number of "yesses" and "noes" and presented these 10 to S while E explained, "These are the words you will see in the sentences to come."

All stimuli appeared on the screen in uppercase letters in white on a dark green field.

Prefacing the experimental trials were 20 practice trials intended to acquaint Ss with the procedure and provide a warm-up. The practice sentences were in blocks of four, giving Ss time for questions after each block. The last three blocks sampled the experimental sentence structures (without duplicating category selection). The first two blocks contained shorter sentences, from two to four categories presented in order of increasing length. This procedure exposed Ss to sentences graduated in difficulty. On the first eight practice trials, Ss received feedback on accuracy from the experimenter (who was cued by the program) and was returned to the misperformed instructions to re-read them. The latter procedure served to stress the accuracy constraint.

Each began with the signal "ready" flashed on the center of the display screen for two seconds. Immediately after the offset of the signal, S saw one of the sentences with its initial letter near the screen margin to the left of center. The sentence remained on the screen until S pressed a button, terminating the trial. If S had not responded after 100 seconds, the sentence disappeared and S was forced to select a response.* A two-second interval separated the trials.

Trials were presented in six blocks of 16, with a rest period of approximately 30 seconds between blocks and a five-minute break after three blocks. E remained in the room with S throughout the session.

Ss were timed in hundredths of a second from sentence onset to the first push of a response button. Response accuracy was recorded at two levels: (a) S's judgment of sentence truth value (indicated by S's choice between the "false" button, "NA," and the "true" buttons, A or B), and (b) given a correct "true" judgment, S's choice between buttons A and B. S's truth judgments were scored by an algorithm which verified each

*There were no cases of latencies this long.

antecedent statement according to its coordinate structure combined with S's truth values on the individual words.

Design

A 4 x 6 x 2 x 2 factorial design was used to assess response time and accuracy. This design collapsed two of the five within-subjects variables -- the nested factor (sentence subtype) with organizational type. The four factors were (a) four numbers of categories, (b) six organizational types, (c) "and-or," and (d) affirmative-negative. Categories and order of sentence presentation were treated as random variables.

Subjects

Forty adults, 16 males and 24 females, participated in the experiment. All were native English speakers. Half were summer students at George Washington University, paid $5.00 for their participation. The rest were white-collar employees of a community service foundation, who donated their time.

Results and Discussion

Before presenting the results obtained in this study, it is important to distinguish between two distinct types of possible outcomes. As was mentioned previously, practically all of the prior research involving the kinds of sentence dimensions employed in this study is of the "main effect" sort; that is, researchers have been concerned with, for example, the difference between "and" and "or" or the number of negations, etc. From the perspective of a document designer, this type of unidimensional study produces a directly meaningful result (e.g., "ands" are easier than "ors"). Furthermore, there are little or no data existing that would suggest that the sentence dimensions are not independent. Thus, one expectation or possible configuration of the results could be that each of the sentence dimensions would have a significant main effect (in an ANOVA sense), and that there would be no significant interactions. This outcome would be ideal for document designers [and equally ideal from a theoretic viewpoint; for example, such a result would allow for an "additive factor" interpretation of component processes (see e.g., Rose, 1978)].

Alternatively, and less satisfying from an applications viewpoint, a second possible outcome would be that the sentence dimensions interact (i.e., the effects of one dimension are different, depending upon levels of other dimensions). Implications of such a pattern of results would be of a different nature than those of the "main effects" pattern. Design recommendations could not be of the sort: "Do not use more than six categories in a conditional sentence." Rather, they would be of the sort: "Do not use more than six categories in hierarchically-organized sentences where the major connective is an or." The reason for raising this fairly obvious point is just that the pattern of results is itself an important

outcome of this study. The pattern of results alone, without any knowledge of the details, will dictate the degree to which design recommendations must be qualified. A complex, interactive pattern, if found, could seriously compromise existing practical "solutions."

As a consequence of these considerations, the first step in the analysis was to run ANOVAs on the entire 4 x 6 x 2 x 2 data set (one for response time and one for proportion correct). If the pattern of results were of the first type (i.e., significant main effects and no significant interactions), then interpretation and presentation of the results would be straightforward. However, the pattern of results showed that a simple interpretation would not be possible. For both the response time and proportion correct analyses, a significant three-way interaction was found among number of categories (C), sentence type (S), and the and-or dimension (A).*

In an attempt to make sense out of the CSA interactions, we decided to conduct separate ANOVAs for each of the major sentence types (strings, single groupings, and "nots" groupings). It was felt that the loss in degrees of freedom due to separating the analyses would be compensated for by clarity of interpretation.** Thus, six ANOVAs, each a 4 x 2 x 2 x 2 design (number of categories by and-or by negation by sentence subtype) were conducted, three each for response time and proportion correct.

---

*The main effect of negation (N) was significant ($p < .05$), furthermore, the dimension did not interact with the other dimensions. In the interest of continuity of presentation, this finding will be discussed in the context of the additional analyses described below.

**The justification for conducting separate analyses on levels of this dimension (as opposed to, for example, the four levels of the number-of-categories dimension) is that the sentence types are basically categorical or nominal variables.

Results of Analysis

Appendix A presents the complete ANOVAs conducted on response time and proportion correct measures for the sentence types. A brief scan of these results suggests that the strategy of analyzing the three types of sentences separately was at least partially successful. Except for "simple grouping response time," all three-way and four-way interactions were nonsignificant ($p > .05$). As was the case in the overall analyses, the general pattern of results was of the "interaction" type; with few exceptions (principally the effect of negations), each of the sentence dimensions interacted significantly with other dimensions.

Given the large number of significant effects, presentation and discussion of these results present a problem. This is exacerbated by the lack of existing theories or models which can address the interactive patterns obtained. As a compromise between potentially overloading the reader with numbers and potentially omitting an important result, we will graphically present the data for each of the sentence types, and will discuss those aspects of the data which appear to be systematically interpretable.

Response Time

Figures 1 to 6 present the results for each sentence type (and subtype) in terms of speed (response time) and accuracy (proportion correct). Each plot shows functions relating number of categories to the "and-or" variable. The remaining sentence dimension (simple negation) is not represented in these plots. This variable tended not to interact with the others. It seemed to have a constant effect: the presence of negative categories uniformly increased response time and decreased accuracy, as shown in Figure 7.

It is clear from Figures 1 to 6 that the relationship between number of categories and the "and-or" dimension varies as a

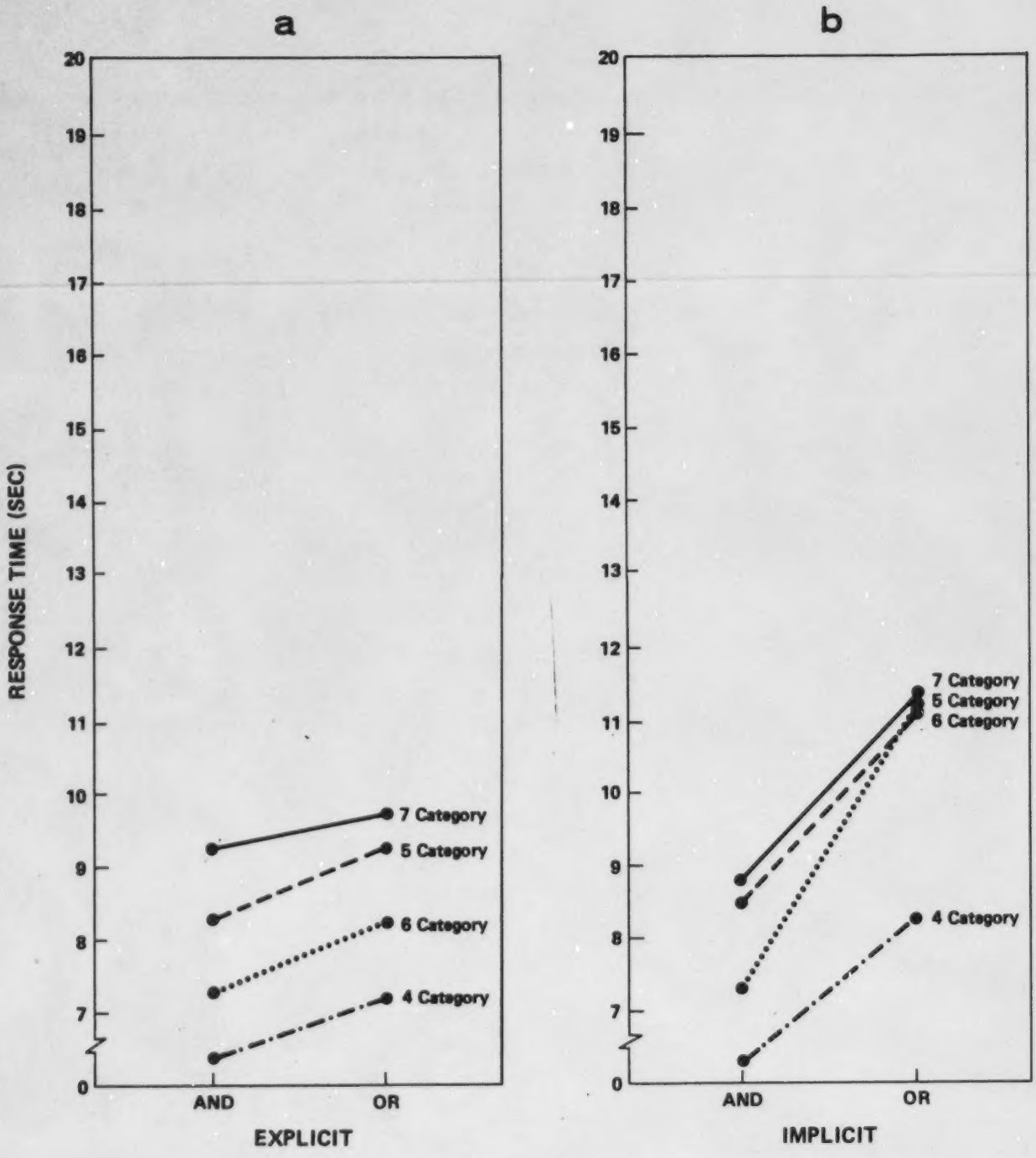

FIGURE 1. Response Time for Strings

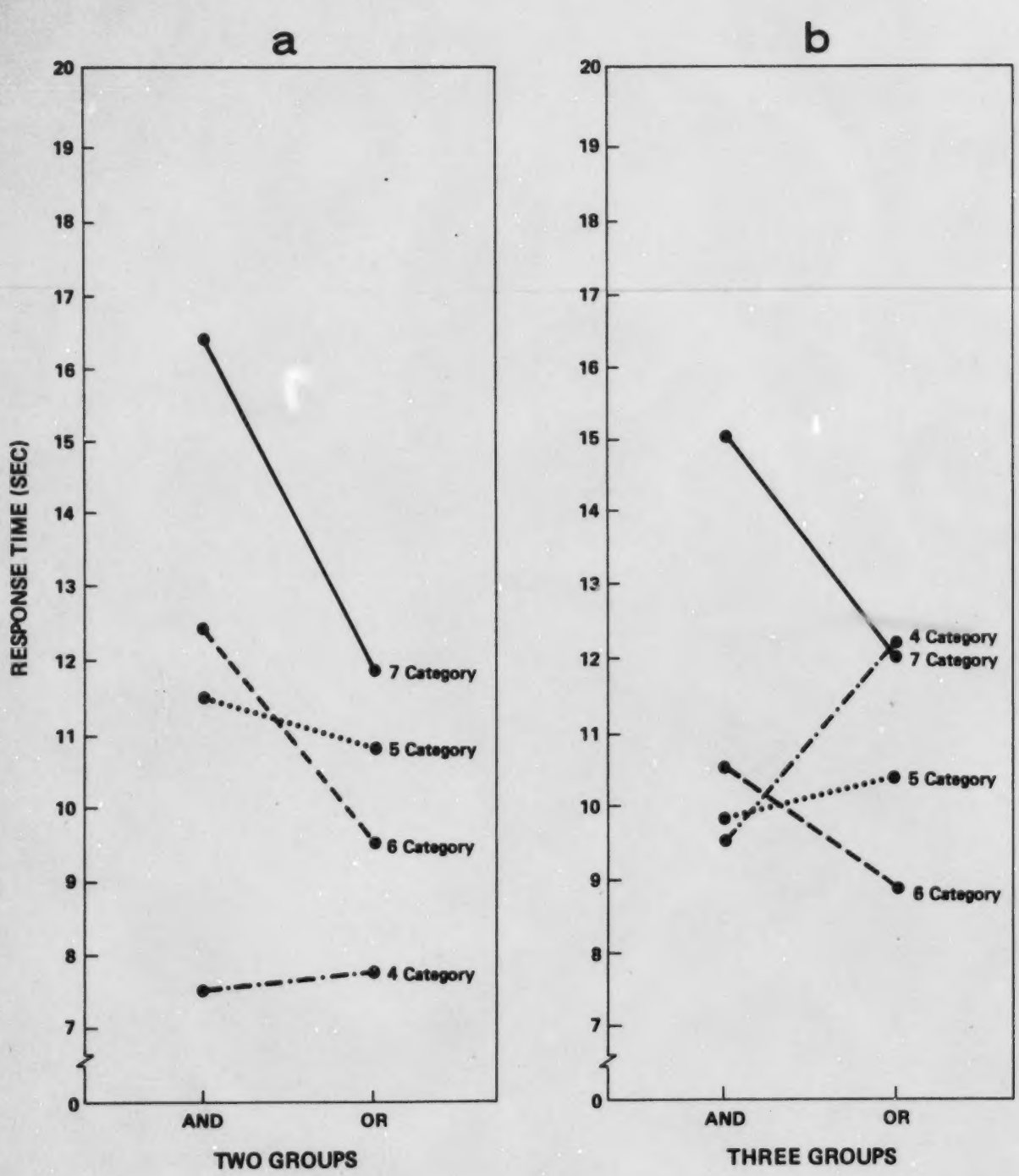

FIGURE 2. Response Time for Simple Groupings

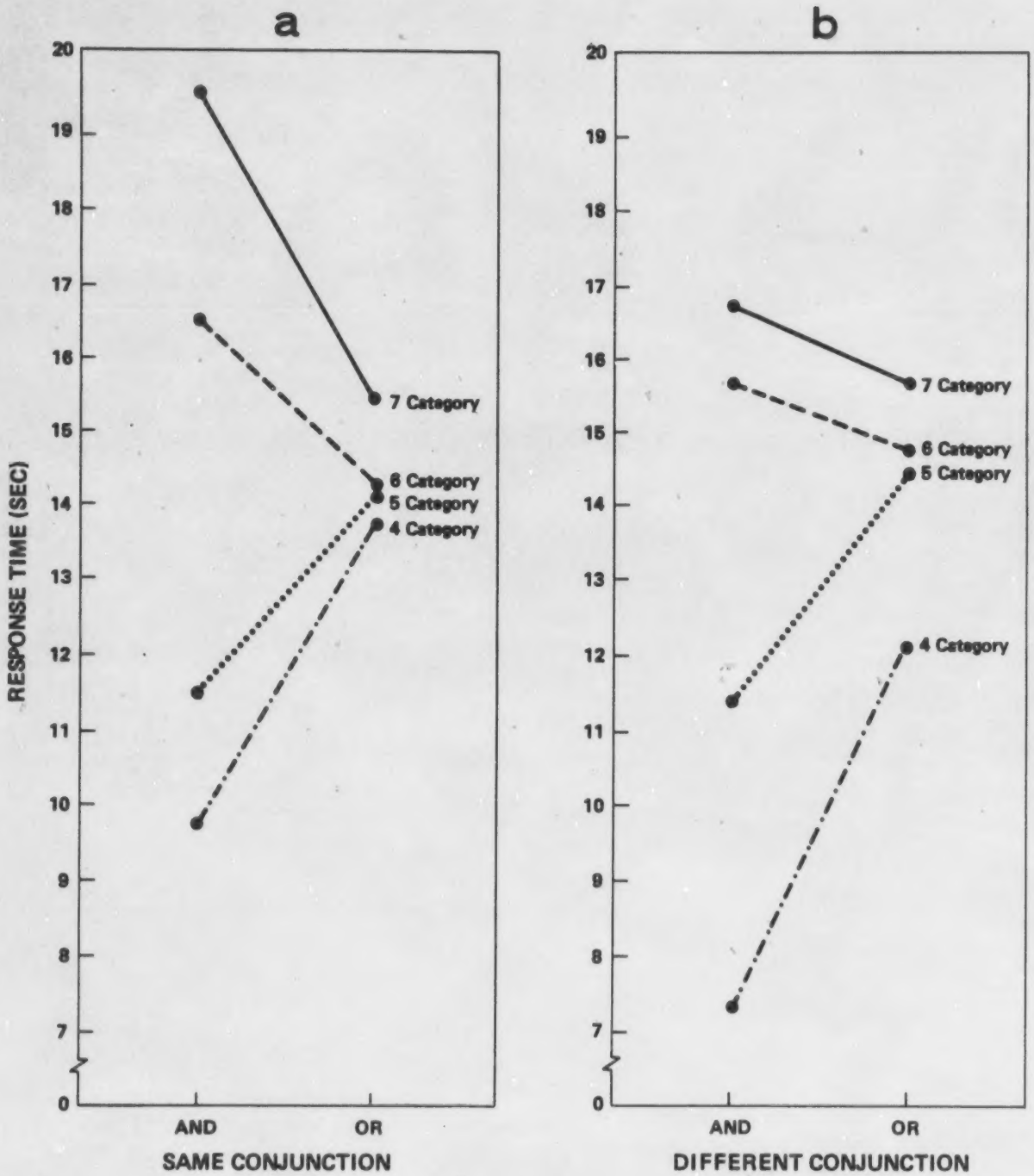

FIGURE 3. Response Time for "Not" Groupings

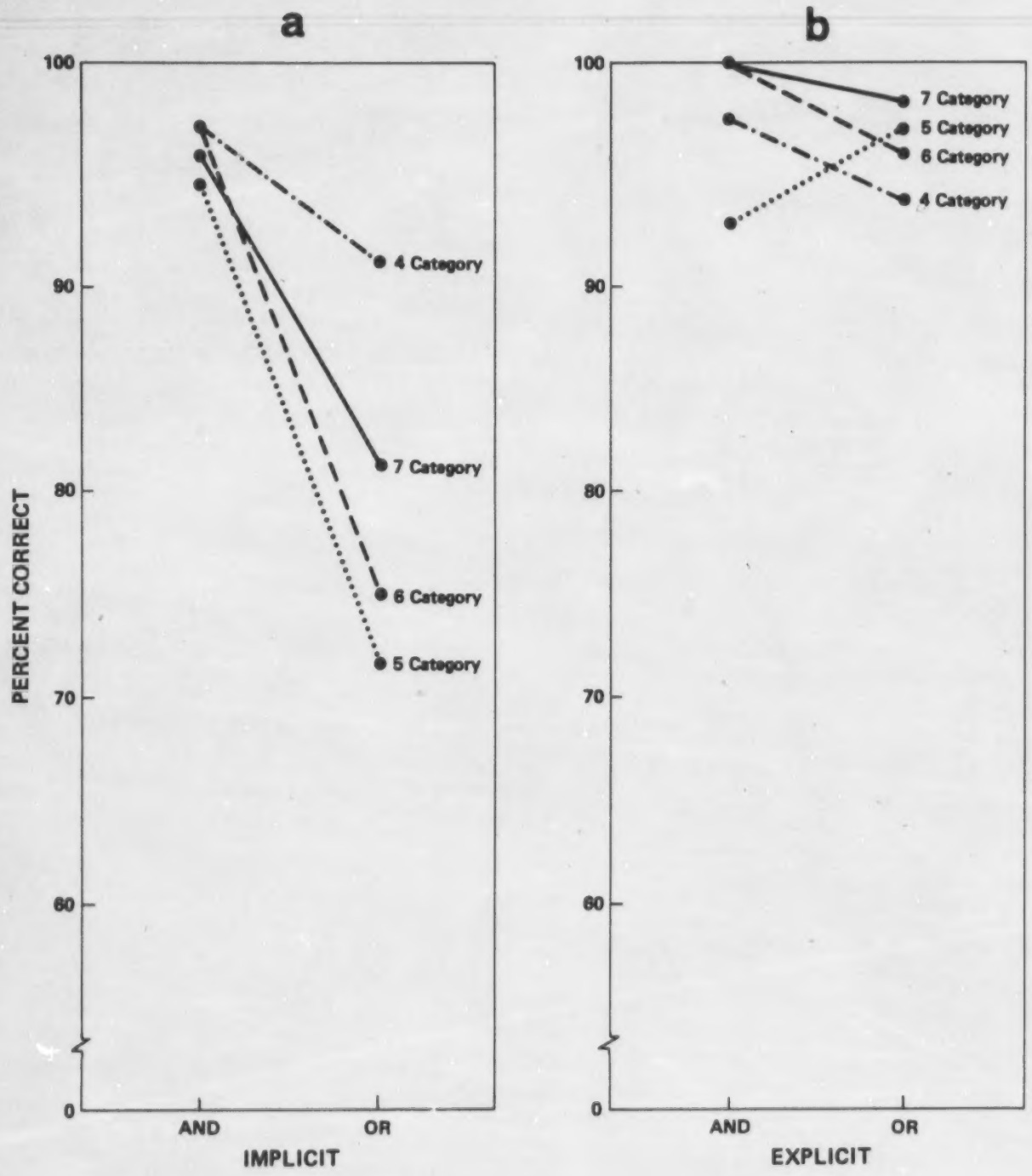

FIGURE 4. Percent Correct for Strings

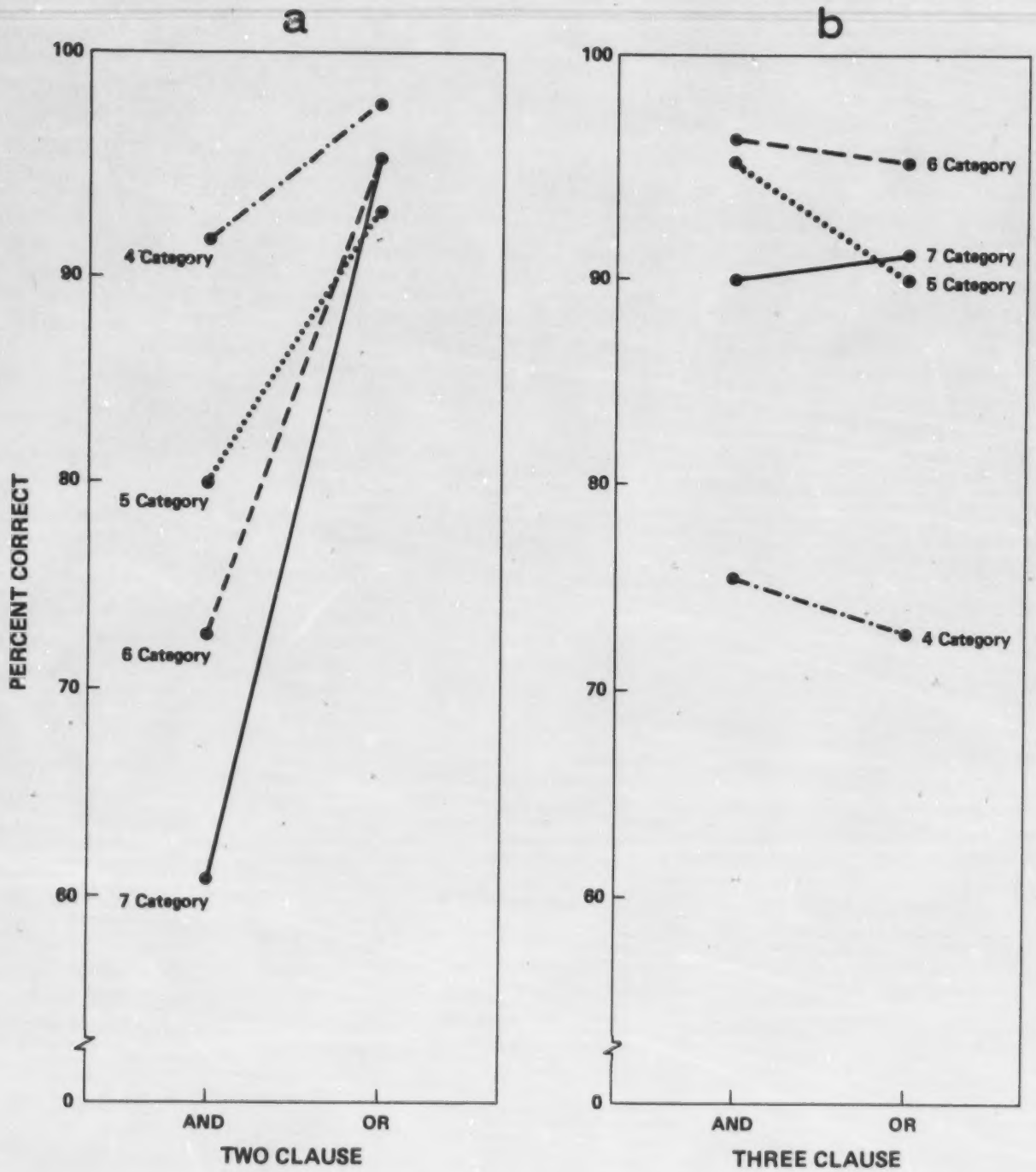

FIGURE 5. Percent Correct for Simple Groupings

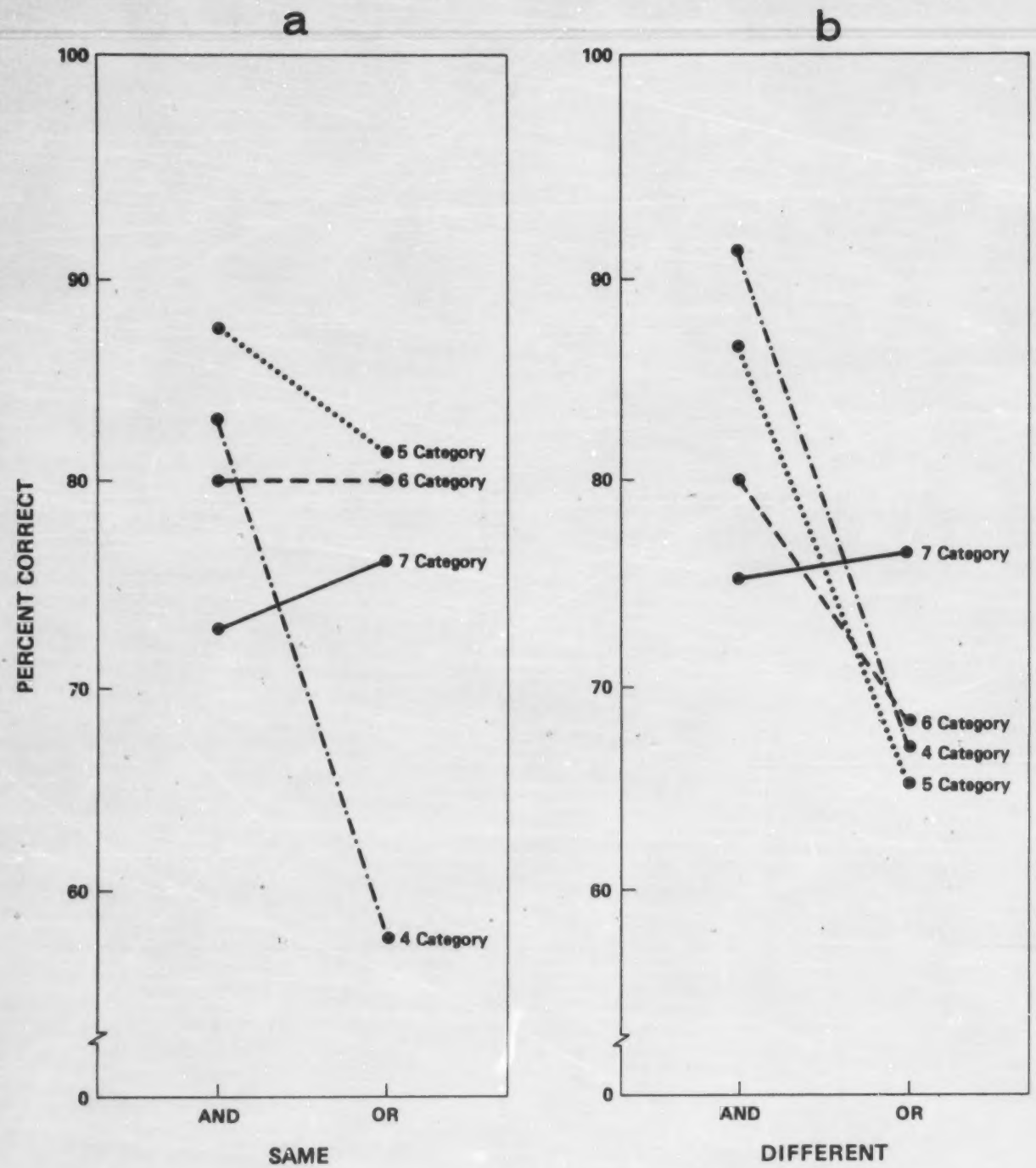

FIGURE 6. Percent Correct for "Not" Groupings

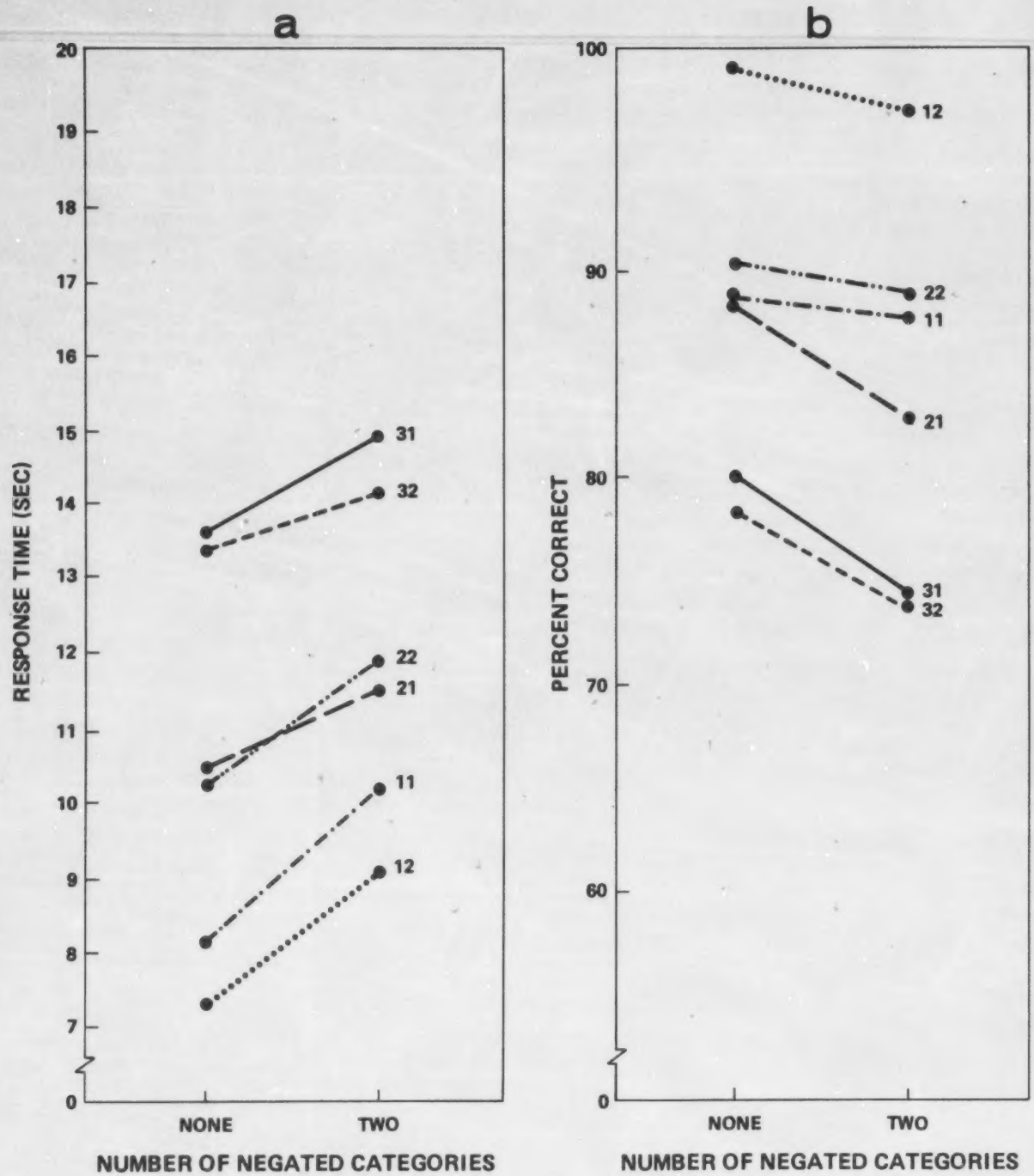

FIGURE 7. Response Times (a) and Percent Correct (b) for Single - Category Negation

function of a sentence type and subtype. The following discussion is an attempt to organize these seemingly disparate findings. We will tentatively develop a model to account for what we believe to be important aspects of the data.

The starting point for this discussion will be the response time plot for explicit strings, shown in Figure 1B. The salient features of this plot are that: there is a uniform increase in response time as a function of number of categories; and there is a uniform increase in time for "or" as compared to "and."

Given the "unambiguous" picture of the explicit sentences, it is tempting to speculate on the underlying processing demands. For example, it is plausible to assume that each additional category adds a constant processing "load." This implies that Ss did not adopt any partial processing strategies. If they had, one would expect no differences in response time as the number of categories increases. [Since each category has an equal probability of being true or false and since the positions of "true" or "false" categories was random, the S strategy of stopping when a "true" category appeared (for "or" sentences) or when a "false" category appeared (for "and" sentences) would have resulted in no differences as a function of the number of categories.] Another way of stating this is that Ss exhaustively perform all of the categorizations in each sentence. Of course, it is possible that, given extensive practice on sentence strings, this strategy would evolve into a more efficient "optional stopping" mode. Nonetheless, Ss in this study did not adopt such a strategy.

Another result that could have processing implications is that the "or" operation has a reasonably constant effect: "or" sentences take about a second longer than "and" sentences. This implies that the "or-effect" is not a function of each individual occurrence of an "or" (or else there would be an increase in the slope of the functions with increasing number

of categories). Rather, it is as if the "or" mode requires subjects to make a single additional categorical decision beyond what they do for "and" sentences. This presumably generalizes to a simple two category case (the response to "A and B" would be faster than to "A or B"); in fact, this is one of the primary empirical results observed in Rose and Cox (1980).

To summarize, a plausible hypothesis is that subjects: (a) make all of the "category" decisions for each sentence, and (b) decide whether the sentence is an "and" or "or" type. If it is an "or" type, it requires an additional time increment to judge the truth or falsity of the entire sentence.

A further clue to what might be happening is the finding that the response times for implicit "and" sentences are virtually identical to explicit "and" sentences (Figure 1a). This implies that subjects were treating all implicit sentences (e.g., "A, B, C, and D") as if the commas were "ands." If this were true, then subjects would require an additional "transformation" if the connective was actually an "or." This explains why the implicit "or" sentences were substantially more difficult than explicit "or" sentences: if the subjects' internal representation of any implicit sentence is as an "and" string, these representations would require an additional transformation to an "or" form.*

Again, a summary is in order regarding our hypothesized processing requirements. Based on a string-type sentences, the indications are that:

---

*At the risk of overspeculating, it is possible to account for the substantial difference between the four-category and higher-number-of-category implicit "or" strings in terms of memory capacity. Suppose subjects store a series of yesses and noes (the individual category decisions). Then they are "hit" with the requirement to reprocess when they discover that the sentence is an "or" sentence. This additional burden might be possible to carry if the load is not already large (i.e., four "bits"); however, if there are more than four bits stored, subjects might be overloaded and be forced to reprocess the entire sentence.

(a) processing load increases directly with number of categories (i.e., there are no "partial processing" strategies employed);

(a) implicit strings are represented as having "and" connectors; hence, implicit "or" strings are more difficult to process than implicit "and" strings (or, more generally, "or" strings are harder than "and" strings).

An explanation of these results is offered by the processing model in Appendix A. This model posits a constant time increment (1) for "or" over "and" and (2) for implicit "or" over explicit "or." The fact that these constants are added for the occurrence of "or" within as well as between groups largely accounts for the pattern of comparisons above.

The data within "and" sentences can be more simply described (Fig. 2b): four categories are faster than five categories, which are faster than six categories, which are substantially faster than seven categories. But the observed ordering of "or" sentences almost reverses that of "and": the four- and five-category sentences take longer than six categories. This complication may be explainable as an effect of semicolons. In constructing the four- and five-category, three-group sentences, we were forced to include semicolons as single-category separators to avoid formally ambiguous sentences.* It is possible that the effect of such "anomalous" semicolons is to increase the time increment incurred by an implicit "or" connective. That is, if subjects are assuming that implicit connectives are "ands," they would be even more prone to make such an assumption if the punctuation is a semicolon.

*Recall that semicolons were not necessary in these sentences but were inserted to equate the means for separating groupings across sentences. The six- and seven-category sentences required semicolons to avoid formal ambiguity.

Turning now to the results for single groupings, we will look first at the response times for three branches (Fig. 2b). These results can be described by a set of simple comparisons between "and" and "or" sentence types:

- four-category "or" takes longer than four-category "and";
- five-category "or" and "and" are approximately equal;
- the difference between "and" and "or" for four categories is larger than the difference for five categories;
- six-category "and" takes longer than six-category "or";
- seven-category "and" takes longer than seven-category "or";
- the difference between "and" and "or" for seven categories is greater than the difference for six categories.

The results obtained for two-branch groupings are shown in Figure 2a. As with the three-branch groupings, these results follow the pattern predicted by the processing model. There are two exceptions. The first is that five-category "ors" took longer than six-category "ors." We have no explanation for this result. The second exception is that for four-category sentences, "ors" were slightly longer than "ands." It could be hypothesized that the "or" decision takes longer when the "or" connects groups (as in the "or" sentences) rather than single categories (as in the "and" sentences). (This hypothesis does not alter the explanation for three-branch groupings.) One final comparison: if we examine the results for two- vs. three-group sentences (Figs. 2a and 2b), these again fit the predictions made by the model.*

The pattern of response time results for the "not" groupings is, in a sense, easily interpretable. It should be recalled that these sentences are basically similar to the three-branch groupings; the major difference is that a "not both" or "not either" was inserted in the former. If we assume that

---

*Also, this comparison lends credence to the notion that our inclusion of a semicolon served both to confuse subjects in the four- and five-category "or" sentences and also generally slowed responses for "and" sentences.

this addition simply makes all decisions harder, then the pattern of results obtained for the single groupings should generalize to the "not" groupings as well. As can be seen (especially for the "SAME" sentences) in Figure 3, this assumption is supported (cf. Fig. 2b and 3a). As with three-branch groupings, the comparisons between "and" and "or" in "not" groupings show that:

- for four categories, "or" takes longer than "and";
- for five categories, "or" still takes longer than "and" but the difference is smaller than for four categories;
- for six categories, "and" takes longer than "or";
- for seven categories, "and" takes longer than "or," and the difference is greater than for six categories.

Accuracy

The results (Figs. 4-6) indicate that in order to have the response time model "fit" accuracy measures, all that must be done is to treat the parameters as if each also reflects a probability of being performed correctly. With certain exceptions (to be discussed below), the accuracy data reflect the speed data almost directly: the longer a sentence takes, the more errors are committed. In fact, the correlation between response time and errors is highly significant [for $n = 48$ (collapsing across all dimensions except negation and sentence type), $r = .92$, $p < .00005$].

A closer look at the errors made gives a good indication of what most subjects were doing. Subjects in this task can make two different types of errors: (1) they can respond "true" when the sentence is actually false; and (2) they can respond "false" when the sentence is actually true. To borrow terms from signal detection theory, we can call the first type a "false alarm error" and the second a "miss." Table 4 shows the proportion of false alarms and misses for the various sentence dimensions. This table shows clearly that subjects have an overwhelming tendency to make errors

TABLE 4. Probability of Errors by Type

Sentence Dimension	P("T"|f)*	P("F"|t)**
A. Number of Categories		
4	.068	.233
5	.060	.200
6	.070	.199
7	.084	.223
B. Sentence Type		
strings implicit	.033	.212
strings explicit	.006	.036
strings	.020	.124
simple groupings: 2-group	.041	.240
simple groupings: 3-group	.055	.157
groupings	.048	.199
"not" grouping: same	.143	.304
"not" grouping: different	.157	.311
"not" grouping	.150	.308
C. And-Or		
and	.063	.254
or	.087	.196
D. Negation		
zero negations	.070	.182
two negations	.071	.244
Overall	.0703	.2130

*Proportion of false trials on which subjects responded "True"

**Proportion of true trials on which subjects responded "False"

that represent misses rather than false alarms. This tendency is true for all partitions of the stimulus set. Although not shown, this same pattern of errors is true for 34 out of 40 subjects (a statistically significant finding, based on a sign test; the binomial probability of 34 out of 40 occurrences by chance = .000004).

This tendency to make "false" responses to factually true sentences has a potentially interesting subjective interpretation. To idealize this tendency, the effect is that when subjects respond "true" they are almost always correct; conversely, whenever they are not sure, they respond with "false." It is as if they were unwilling to admit that they belonged to a complex category. Again, to borrow an analogy from Signal Detection Theory, the subject's criterion was radically shifted in a conservative direction. Another indication of what subjects were doing comes from the observation that the mean response time for all trials on which subjects responded "false" (56.9% of all trials) was 11.01 sec.; for trials on which they responded "true" (43.1%), mean response time was 11.54 sec. Thus, it appears that subjects were not "fast-guessing"; rather, they attempted to solve all problems.

Specific sources of difficulty (beyond those implied in the previous discussion of response time) and the exceptions to the speed-accuracy correspondence can be hypothesized from an examination of the accuracy plots (Figs. 4-6). The first feature to note is for explicit strings, where practically no errors were made. Harking back to our original conceptualization of the task, it will be recalled that subjects must perform three processes: they must extract the logical rule, evaluate the conditions, and decide on a response. In the explicit string condition, the "extraction of the rule" is simplified by making the rule overt. Thus, the only source of error (outside of "nonexperimental" influences, such as a finger slipping or a total misunderstanding of the meanings

of "and" and "or") would be due to misclassification at the individual category level. The results from this condition confirm that subjects simply did not make errors at this level.

Another feature to note is the apparently anomalous results for four-category, three-group hierarchies. Our guess is that for "or" sentences this is a function of the subjects' misinterpreting of the semicolons as meaning "and." This misinterpretation persists because there are no explicit "ors" at four categories to connect between semicolons (as there are at the higher numbers of categories). Why subjects did relatively poorly on four-category "ands" is not clear. This same explanation applies to four-category "ors" for phrasal "not" sentences.

Summary and Implications

To recapitulate the major results, we can make the following points:

1. The pattern of results, for both speed and accuracy measures, was primarily of the interactive type. No general interpretation of variable effects could be made without qualifying them.
2. Each of the major sentence dimensions (number of categories, sentence type, and-or, and negation) seemed to contribute to the overall difficulty of the sentences.
3. Subjects did not appear to adopt any partial processing strategies.
4. When a logical connector was not explicitly stated, subjects acted as if it was an "and" rather than an "or."
5. Semicolons seemed to exacerbate processing difficulties.
6. Speed and accuracy were related.
7. Subjects were overwhelmingly prone to make one type of error rather than another: many more errors of the "miss" sort were made than of the "false alarm" sort.

From a theoretical perspective, these results have mixed implications. Our interpretations of the data are consistent with most theoretical representations of a sentence verification process in their most general form: the sentences are encoded and transformed to some more simple representation, truth values are ascertained for individual categories, a logical rule is extracted, and a decision is made by comparing the stored representation to previously-learned information. Perhaps the most conspicuous finding regarding these theories is that subjects seem to be representing strings of categories as if they were connected by "ands" rather than "ors." Thus, the typical finding of a conjunctive-disjunctive difference might be due to some fundamental encoding bias. Further support for this notion comes from the observation that, for the explicit "or" strings, there is a uniform increase in response time, regardless of the number of "ors" in a sentence.

On the other hand, this study says little about how people extract logical rules from the sentences.* In fact, the data are consistent with a processing system that does not extract such rules. That is, the pattern of results could be generated by a continuous, serial, left-to-right updating process, which determines a "running" truth value that changes whenever a new category is evaluated.

At a more basic level, these results reveal the limitations of all existing theories. Those theories simply do not have sufficient mechanisms to account for the obviously complex relationships involved when people process multicategory concepts. It should also be noted that we did not include other, "real-world" dimensions that would further complicate the issues. For example, we chose not to include "difficulty of categorization" as a variable. In most forms and instructions, the individual category decisions are themselves difficult

*"Rules" here refers to the truth functional relations of logic and not to syntactic or semantic structure.

(e.g., "if you earned between $10,000 and $25,000 in three of the last six years" or "if you were a student in any three years between 1960 and 1968"). We suspect that such a variable would have radical effects on most of the sentence types used in this study.

Practical Implications

One of the motivations for this study was to see if we could find "thresholds" -- levels of variables which clearly indicated that subjects were having difficulties. The identification of such "problem" sentences would serve two purposes: first, we could make specific recommendations to writers and document designers as to which sentence constructions should and should not be used. Second, these sentences would provide us with a stimulus pool for which we could generate and evaluate different presentation formats (e.g., algorithms, decision trees, etc.).

It is fairly straightforward (albeit entirely arbitrary) to identify specific combinations of variables which resulted in good or poor performance. At a gross level, and using a criterion of approximately 90% correct as a cutoff, we can make the following recommendations:

A. "Good" sentences: those which would be acceptable for general use:

- implicit strings, where the connective is an "and";
- explicit strings, regardless of the connective or the number of categories;
- two-group hierarchical sentences, where the major connective is an "or";
- three-group hierarchical sentences with more than four categories (or, better still, without semicolons).*

*Semicolons can be eliminated without reverting to ambiguity by repeating the sentence subject to clarify groupings: "You are A, B, and C, or you are D, E, and F" instead of "You are A, B, and C; or D, E, and F."

B. "Poor" sentences: those which clearly should be rewritten or redesigned:

- any sentence containing a "not" grouping;
- implicit strings with an "or" connective;
- two-group hierarchies with more than four categories, where the major connective is an "and";*
- three-group hierarchies of less than six categories (unless groupings are clarified by other means than semicolons).

Of course, a different criterion (such as one which included response time as a criterion) would change the set of recommendations. Likewise, any additional information the document designer might possess could be potentially useful (e.g., the relative cost of "misses" vs. "false alarms" with regard to document processing; the length of the form and subjects' desires regarding the investment of time needed to complete it, etc.). The safest recommendation to make at this point in our research program is that all complex instructions should be experimentally evaluated prior to inclusion in any document designed for public use.

---

*Recall the explanation behind this recommendation: Any structure containing one or more implicit "or" series ("A, B, or C") will be difficult. Two-group "and" sentences of five categories or more are just such structures.

REFERENCES

Anderson, J.R. *Language, memory and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. & Reder, L. Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 664-681.

Bever, T.G. The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley, 1970.

Bourne, L.E. Learning and utilization of conceptual ideas. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory.* New York: Wiley, 1967.

Braine, M.D.S. Children's first word combinations. *Monographs of the Society for Research in Child Development,* 1976, *41* (1, Serial No. 164).

Braine, M.D.S. On the relation between the natural logic of reasoning and standard logic. *Psychological Review,* 1978, *85,* 1-21.

Bruner, J.S., Goodnow, J.J.,& Austin, G.A. *A study of thinking.* New York: Wiley, 1956.

Carpenter, P. & Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review,* 1975, *82,* 45-73.

Charrow, R. & Charrow, V. Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review,* 1979, *79,* 1306-1374.

Charrow, V. & Charrow, R. *The comprehension of standard jury instructions: A psycholinguistic approach.* Arlington, Va: Center for Applied Linguistics, 1978.

Chase, W.G. & Clark, H.H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972.

Clark, H.H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, *76,* 387-404.

Clark, H.H. Difficulties people have in answering the question "Where is it?" *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 265-277.

Clark H.H. Semantics and comprehension. In T.A. Sebeck (Ed.), *Current trends in linguistics* (Vol. 1). The Hague: Mouton, 1973.

Clark, H.H. & Card, S.K. The role of semantics in remembering comparative sentences. *Journal of Experimental Psychology*, 1969, *82*, 545-553.

Clark, H.H. & Chase, W.G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, *3*, 472-517.

Collins, A.M. & Quillian, M.R. Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 432-438.

Fillmore, C. Case for case. In E. Bach & R. Harms, (Eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, & Winston, 1968.

Fodor, J.A., Bever, T.G., & Garrett, M.F. *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill, 1974.

Gough, P.B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning & Verbal Behavior*, 1965, *5*, 107-111.

Gough, P.B. The verification of sentences: The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 492-496.

Greene, J.M. The semantic function of negatives and passives. *British Journal of Psychology*, 1970, *22*, 14-27.

Griggs, R.A. & Shea, S.L. Integrating verbal quantitative information in linear orderings. *Memory and Cognition*, 1977, *5*, 287-291.

Haygood, R.C. & Bourne, L.E. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, *72*, 175-195.

Horabin, I. & Lewis, B. *Algorithms*. Charles Town, W. Va.: Ivan Horabin, 1974.

Hunt, E.B. & Hovland, C.I. Order of consideration of different types of concepts. *Journal of Experimental Psychology*, 1960, *59*, 220-225.

Johnson, N.F. The psychological reality of phrase structure rules. *Journal of Verbal Learning & Verbal Behavior*, 1965, *4*, 469-475.

Johnson, N.F. Sequential verbal behavior. In T.R. Dixon & D.L. Horton (Eds.), *Verbal Behavior and General Behavior Theory*. Englewood Cliffs: Prentice Hall, 1968.

Johnson-Laird, P.N. The interpretation of quantified sentences. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in Psycholinguistics*. Amsterdam: North-Holland, 1970.

Johnson-Laird, P.N. Linguistic complexity and insight into a deductive problem. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in Psycholinguistics*. Amsterdam: North-Holland, 1970.

Johnson-Laird, P.N. Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and processes in children and adults*. Hillsdale, N.J.: Erlbaum, 1975.

Johnson-Laird, P.N. & Tagart, J. How implication is understood. *American Journal of Psychology*, 1969, *82*, 367-373.

Jones, S. Instructions, self-instructions and performances. *Quarterly Journal of Experimental Psychology*, 1968, *20*, 74-78.

Just, M.A. Comprehending quantified sentences: The relation between sentence-pictures and semantic memory verification. *Cognitive Psychology*, 1974, *6*, 216-236.

Just, M. & Carpenter, P. Comprehension of negatives with quantification. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 244-253.

Just, M. & Clark, H.H. Drawing inferences from the presuppositions and implications of affirmative and negative sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 21-31.

Katz, J.J. *Semantic theory*. New York: Harper and Row, 1972.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Lakoff, G. Linguistics and natural logic. *Synthese*, 1970, *22*, 151-271.

Legrenzi, P. Relations between language and reasoning about deductive rules. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North-Holland, 1970.

Loftus, E.G. Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 1973, *97*, 70-74.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning & Verbal Behavior*, 1963, *2*, 346-351.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81-97.

Miller, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, *17*, 748-762.

Miller, G.A. & McKean, K. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, 1964, *16*, 297-308.

Morris, J.O. *Make yourself clear!* New York: McGraw-Hill, 1972.

Neisser, U. & Weene, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, *64*, 640-645.

Perfetti, C.A. Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 719-724.

Quine, W.V.O. *Elementary logic.* New York: Harper & Row, 1965.

Rips, L.J., Smith, E.E., & Shoben, E.J. Semantic composition in sentence verification. *Journal of Verbal Learning & Verbal Behavior*, 1978, *17*, 375-402.

Rohrman, N.L. The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning & Verbal Behavior*, 1968, *7*, 904-912.

Rose, A.M. *An information-processing approach to performance assessment.* (Final report). Washington, D.C.: American Institutes for Research, November 1978.

Rose, A.M. & Cox, A. *Following instructions.* (Final report). Washington, D.C.: American Institutes for Research, 1980 (in preparation).

Sherman, M. Adjectival negation and the comprehension of multiply-negated sentences. *Journal of Verbal Learning & Verbal Behavior*, 1976, *15*, 143-157.

Sherman, M.A. Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 76-84.

Slobin, D.I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 219-227.

Smith, E.E., Shoben, E.J., & Rips, L.J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychology Review*, 1974, *81*, 214-241.

Trabasso, T. Mental operations in language comprehension. In J.B. Carroll & R.D. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: V.H. Winston, 1972.

Trabasso, T., Rollins, H., & Shaughnessey, E. Storage and verification shapes in processing concepts. *Cognitive Psychology*, 1971, *2*, 231-289.

Wales, R.J. & Grieve, R. What is so difficult about negation? *Perception and Psychophysics*, 1969, *6*, 327-332.

Wason, P.C. *Psychological aspects of negation*. London: Communications Research Centre, University College, 1962.

Wason, P.C. The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 7-11.

Wason, P.C. The processing of positive and negative information. *Quantitative Journal of Experimental Psychology*, 1959, *11*, 92-107.

Wason, P.C. Psychological aspects of inference. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North-Holland, 1970.

Wason, P.C. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, *52*, 133-142.

Wason, P.C. & Johnson-Laird, P.N. *Psychology of reasoning: Structure and content*. Cambridge, Mass.: Harvard University Press, 1972.

Wason, P.C. & Jones, S. Negatives: Denotation and connotation. *British Journal of Psychology*, 1963, *54*, 299-307.

Yngve, V.H. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 1960, *104*, 444-466.

Appendix A

ANOVA Tables for Each of the Experimental Sentence Types

Strings Reaction Time

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	1447.1	3	482.4	23.6	.000
Error	2391.7	117	20.4		
Surface Structure (S)	265.2	1	265.2	9.6	.004
Error	1073.5	39	27.5		
CS	46.4	3	15.5	.8	.478
Error	2169.0	117	18.5		
And / Or (A)	1008.4	1	1008.4	39.8	.000
Error	987.8	39	25.3		
CA	49.0	3	16.3	1.0	.419
Error	2009.0	117	17.2		
SA	308.0	1	308.0	18.2	.000
Error	662.0	39	17.0		
CSA	30.5	3	10.2	.6	.624
Error	2024.3	117	17.3		
Negation (N)	1125.5	1	1125.5	58.1	.000
Error	755.4	39	19.4		
CN	13.2	3	4.4	.3	.855
Error	1979.5	117	16.9		
SN	11.5	1	11.5	.9	.348
Error	496.5	39	12.7		
CSN	79.1	3	26.4	1.4	.247
Error	2203.8	117	18.8		
AN	.7	1	.7	.0	.861
Error	845.3	39	21.7		
CAN	5.8	3	1.9	.1	.944
Error	1793.2	117	15.3		
SAN	.1	1	.1	.0	.945
Error	1036.0	39	26.6		
CSAN	29.2	3	9.7	.5	.679
Error	2255.4	117	19.3		

Strings Percent Correct

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	.32	3	.11	1.69	.172
Error	7.40	117	.06		
Surface Structures (S)	3.10	1	3.10	28.09	.000
Error	4.31	39	.11		
CS	.90	3	.30	6.89	.000
Error	5.07	117	.04		
And / Or (A)	3.10	1	3.10	20.39	.000
Error	5.93	39	.15		
CA	.38	3	.13	2.15	.098
Error	6.96	117	.06		
SA	1.58	1	1.58	9.95	.003
Error	6.20	39	.16		
CSA	.41	3	.14	2.62	.054
Error	6.18	117	.05		
Negation (N)	.06	1	.06	1.44	.238
Error	1.72	39	.04		
CN	.07	3	.02	.61	.608
Error	4.52	117	.04		
SN	.00	1	.00	.02	.897
Error	1.78	39	.05		
CSN	.17	3	.06	1.51	.216
Error	4.42	117	.04		
AN	.28	1	.28	5.18	.028
Error	2.12	39	.05		
CAN	.03	3	.01	.22	.885
Error	4.94	117	.04		
SAN	.06	1	.06	1.34	.254
Error	1.84	39	.05		
CSAN	.31	3	.10	2.89	.038
Error	4.16	117	.04		

Simple Groupings Reaction Time

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	3792.7	3	1264.2	27.59	.000
Error	5361.5	117	45.8		
Surface Structures (S)	.9	1	.9	.02	.881
Error	1560.4	39	40.0		
CS	1060.1	3	353.4	8.40	.000
Error	4921.8	117	42.1		
And / Or (A)	449.3	1	449.3	10.39	.003
Error	1686.3	39	43.2		
CA	1276.5	3	425.5	10.07	.000
Error	4943.5	117	42.3		
SA	231.9	1	231.9	7.63	.009
Error	1185.0	39	30.4		
CSA	21.3	3	7.1	.25	.862
Error	3334.5	117	28.5		
Negation (N)	644.5	1	644.5	19.51	.000
Error	1288.2	39	33.0		
CN	140.1	3	46.7	1.21	.311
Error	4527.6	117	38.7		
SN	27.7	1	27.7	.55	.464
Error	1977.2	39	50.7		
CSN	189.3	3	63.1	1.66	.179
Error	4442.8	117	38.0		
AN	202.7	1	202.7	8.02	.007
Error	985.5	39	25.3		
CAN	376.5	3	125.5	4.09	.008
Error	3593.0	117	30.7		
SAN	1.8	1	1.8	.08	.775
Error	863.5	39	22.1		
CSAN	66.8	3	22.3	.74	.531
Error	3531.3	117	30.2		

Simple Groupings Percent Correct

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	.62	3	.21	2.46	.066
Error	9.85	117	.08		
Surface Structures (S)	.41	1	.41	6.47	.015
Error	2.49	39	.06		
CS	4.11	3	1.38	15.13	.000
Error	10.60	117	.09		
And / Or (A)	1.58	1	1.58	10.16	.003
Error	6.07	39	.16		
CA	1.86	3	.62	6.83	.000
Error	10.61	117	.09		
SA	3.94	1	3.94	36.41	.000
Error	4.22	39	.11		
CSA	.33	3	.11	1.15	.334
Error	11.14	117	.10		
Negation (N)	.28	1	.28	3.67	.063
Error	3.00	39	.08		
CN	.37	3	.12	1.45	.232
Error	9.97	117	.09		
SN	.06	1	.06	.55	.462
Error	4.47	39	.11		
CSN	.06	3	.02	.22	.883
Error	11.53	117	.10		
AN	.04	1	.04	.33	.568
Error	4.49	39	.12		
CAN	.03	3	.01	.08	.973
Error	14.07	117	.12		
SAN	.13	1	.13	1.32	.257
Error	3.90	39	.10		
CSAN	.33	3	.11	1.33	.267
Error	9.76	117	.08		

"Not" Groupings Reaction Time

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	5942.7	3	1980.9	37.58	.000
Error	6166.8	117	52.7		
Surface Structures (S)	94.4	1	94.4	2.77	.104
Error	1330.0	39	34.1		
CS	108.5	3	36.2	.61	.611
Error	6954.2	117	59.4		
And / Or (A)	70.0	1	70.0	1.74	.195
Error	1572.8	39	40.3		
CA	2248.0	3	749.3	15.15	.000
Error	5787.9	117	49.5		
SA	62.5	1	62.5	.58	.452
Error	4219.5	39	108.2		
CSA	185.7	3	61.9	1.39	.249
Error	5206.3	117	44.5		
Negation (N)	330.9	1	330.9	3.80	.059
Error	3397.2	39	87.1		
CN	226.0	3	75.3	1.57	.201
Error	5617.6	117	48.0		
SN	33.7	1	33.7	.79	.378
Error	1655.9	39	42.5		
CSN	46.1	3	15.4	.27	.844
Error	6552.2	117	56.0		
AN	13.8	1	13.8	.34	.565
Error	1597.9	39	41.0		
CAN	42.6	3	14.2	.30	.822
Error	5469.8	117	46.8		
SAN	120.6	1	120.6	2.47	.124
Error	1906.8	39	48.9		
CSAN	40.9	3	13.6	.40	.755

"Not" Groupings Percent Correct

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROBABILITY
Number of Categories (C)	.58	3	.19	1.00	.396
Error	22.55	117	.19		
Surface Structures (S)	.03	1	.03	.15	.699
Error	7.22	39	.19		
CS	1.55	3	.52	3.99	.010
Error	15.20	117	.13		
And / Or (A)	3.40	1	3.40	10.33	.003
Error	12.85	39	.33		
CA	3.17	3	1.06	6.15	.001
Error	20.08	117	.17		
SA	.38	1	.38	1.26	.269
Error	11.75	39	.30		
CSA	.34	3	.11	.74	.532
Error	18.03	117	.15		
Negation (N)	.80	1	.80	4.57	.039
Error	6.83	39	.18		
CN	.92	3	.31	1.75	.160
Error	20.46	117	.17		
SN	.01	1	.01	.07	.789
Error	6.74	39	.17		
CSN	.03	3	.01	.06	.979
Error	19.22	117	.16		
AN	.31	1	.31	2.35	.133
Error	5.19	39	.13		
CAN	.64	3	.21	1.23	.301
Error	20.36	117	.17		
SAN	.05	1	.05	.48	.493
Error	4.08	39	.10		
CSAN	.06	3	.02	.13	.944
Error	17.32	117	.15		

Appendix B

An Information-Processing Model

Appendix B

An Information-Processing Model

This Appendix presents the development of an information processing model designed to account for the principal empirical results obtained in the study. This model allows us to conceptualize the sentence verification task as a series of component operations (or processes) that combine linearly to produce individual sentence response time. While the model is highly speculative, it does serve two useful functions: first, it helps to organize and explain the results; and second, it can be used to generate predictions of difficulty for sentence types not included in the present study.

As discussed in the Results section, the data from the explicit strings showed that response time increased directly with the number of categories in the string, and there was an approximately constant difference between "ands" and "ors" across number of categories. Furthermore, a comparison between explicit strings and implicit strings showed that while no additional time was needed for implicit "and" strings (as compared to explicit "and" strings), implicit "ors" added a processing burden.

We can account for these results by hypothesizing four "operations" or individual processes, each of which takes a finite (but measurable) amount of time. To simplify the discussion that follows, we will identify each of these operations by a letter. The four operations are as follows:

a = the time necessary for an individual category decision (implied by the constant difference between number of categories);

b = the additional time necessary for "or" sentences (implied by the constant "and - or" difference);

c = the additional time necessary when an implicit connective is an "or" rather than an "and" (implied by the time increase of implicit "ors" as compared to explicit "ors");

d = the time it takes to judge the truth or falsity of a sentence.

We will also make two processing assumptions. First, subjects exhaustively perform all the categorizations in each sentence. Second, they exhaustively perform all the truth value decisions required in each sentence. To illustrate this second assumption, consider a two-branch grouped sentence of the form"A or B, and C or D." We assume that the subjects make three decisions: (1) the truth or falsity of the "A or B" clause; (2) the truth or falsity of the "C or D" clause; and (3) the truth or falsity of the entire sentence [(A or B) *and* (C or D)].

We can represent each sentence in the study using this notation. For example, a four-category explicit "and" string ("A and B and C and D") would require the following operations:

- four "a" category decisions (one each for A, B, C, and D);
- one "d" judgment (for the truth value of the entire sentence).

Each addition of a category would require an additional "a" operation. Thus, explicit "and" sentences can be represented as follows:

four-category explicit "and" = 4a + d
five-category explicit "and" = 5a + d
six-category explicit "and" = 6a + d
seven category explicit "and" = 7a + d

In contrast, we hypothesize that explicit "or" sentences require an additional "b" operation:

four-category explicit "or" = 4a + b + d
five-category explicit "or" = 5a + b + d
six-category explicit "or" = 6a + b + d
seven-category explicit "or" = 7a + b + d

We assume that implicit "and" strings have the same processing requirements as explicit "and" strings. However, implicit "or" strings require a "c" operation:

four-category implicit "or" = 4a + b + c + d
five-category implicit "or" = 5a + b + c + d
six-category implicit "or" = 6a + b + c + d
seven-category implicit "or" = 7a + b + c + d

For the hierarchically-organized sentences, our processing assumption requires that subjects process (i.e., decide on the truth or falsity of) each group individually. Consider the four-category, two-group "and" sentence "A or B; and C or D." This sentence requires four "a" operations (as before). In addition, each group requires a "b" operation (for the explicit "or" connective) and a "d" operation (to judge the truth or falsity of the clause). Finally, the entire sentence requires a "d" operation (to judge the truth or falsity of the sentence). In summary, the representation of the sentence is:

four-category, two-group "and" = 4a + 2(b+d) + d
= 4a + 2b + 3d

Similarly, five-category, two-group "and" (A or B; and C, D, or E) sentences require five "a" operations; the first clause requires an additional "c" operation and a "d" operation. The resulting representation is :

5a + (b+d) + (b+c+d) + d = 5a + 2b + c + 3d

The other two-clause "and" sentences are represented as follows:

A, B, or C; and D, E, or F = 6a + 2(b+c+d) + d
= 6a + 2b + 2c + 3d

A, B, or C; and D, E, F, or G = 7a + 2(b+c+d)
= 7a + 2b + 2c + 3d

Two-clause "or" sentences are represented as follows:

A and B; or C and D = 4a + 2(d) + b + d
= 4a + b + 3d

A and B; or C, D, and E = 5a + 2(d) + b + d
= 5a + b + 3d

A, B, and C; or D, E, and F = 6a + 2(d) + b + d
= 6a + b + 3d

A, B, and C; or D, E, F, and G = 7a + 2(d) + b + d
= 7a + b + 3d

Using this notation, the following tables show how each of the three-clause grouped sentences would be represented:

	"and" sentences	"or" sentences
four-category	A; B; and C or D	A; B; or C and D
	4a + (b+d) + d = 4a + b + 2d	4a + c + (b+d) + d = 4a + b + c + 2d
five-category	A; and B or C; and D or E	A; or B and C; or D and E
	5a + 2(b+d) + d = 5a + 2b + 3d	5a + 2(b+d) + d = 5a + 2b + 3d
six-category	A or B; and C or D; or F	A and B; or C and D; or E and F
	6a + 3(b+d) + d = 6a + 3b + 4d	6a + 3d + b + d = 6a + b + 4d
seven-category	A or B; and C or D; and E, F, or G	A and B; or C and D; or E, F, and G
	7a + 3(b+d) + c + d = 7a + 3b + c + 4d	7a + 3d + b + d = 7a + b + 4d

These sentence representations can serve to predict the relative speed with which subjects process each sentence type. For example, for the three-branch grouped sentence above, the prediction is that four-category "ors" should take longer than four-category "ands":

4a + b + c + 2d > 4a + b + 2d.

We can make several such predictions. Thus, for the explicit strings, we can obviously "predict" the results for response time in all particulars, since these results were used to generate the model's parameters. With other sentence types, we can theoretically predict each of the functions shown in Figures 1 through 6. As an example, the following predictions can be made regarding the three-branch grouped sentences:

- four-category "or" will take longer than four-category "and";

- five-category "or" and "and" will be approximately equal;
- the difference between "and" and "or" for four categories will be larger than the difference for five categories;
- six-category "and" will take longer than six-category "or";
- seven-category "and" will take longer than seven-category "or";
- the difference between "and" and "or" for seven categories will be greater than the difference for six categories.

An examination of the results (Figure 2b) shows that all of the above predictions were borne out. Of course, these are not all of the predictions that could have been generated, even within this sentence type. In particular, the above predictions are for relative ordering of sentence types, not for absolute values. In order to predict, we would have to hypothesize about the relative durations of the operations. For example, if we assume that $c > b + d$, and $b > a + d$, it would be possible to exactly match the order of the sentences. Likewise, if we incorporate the effect of semicolons on processing time (as discussed in the text), more refined predictions could be made.

Similar comparisons of the predicted and actual results for other sentence types are, for the most part, supportive of the model. However, it must be noted that the characterization of the results in terms of a set of parameters and processing assumptions is primarily a heuristic device for summarizing many different conditions and variations. As such, our presentation still might miss some important effects and/or overemphasize others. Furthermore, this model must be considered as informal in several senses. In order to "qualify" as a formal model (in the sense of being testable), certain additional assumptions must be made. For example, we have treated the combination of parameters as if they were additive. Also, with the exception of the "c" and "b" parameters, we have assumed that parameters have the same values in different sentences (i.e., that the values are unconditional or independent).

Furthermore, we have not provided numerical estimates for the relative or absolute values of these parameters.